In compliance with the
Canadian Privacy Legislation
some supporting forms
may have been removed from
this dissertation.

While these forms may be included
in the document page count,
their removal does not represent
any loss of content from the dissertation.

University of Alberta

Intentions, Conventions, Rules, and Reality

by

Brian Patrick Leahy

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Master of Arts

Department of Philosophy

Edmonton, Alberta
Fall 2003

National Library
of Canada

Acquisitions and
Bibliographic Services

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale
du Canada

Acquisisitons et
services bibliographiques

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 0-612-87580-6
Our file Notre référence
ISBN: 0-612-87580-6

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou aturement reproduits sans son autorisation.

Canada

University of Alberta

Library Release Form

Name of Author: *Brian Patrick Leahy*

Title of Thesis: *Intentions, Conventions, Rules, and Reality*

Degree: *Master of Arts*

Year this Degree Granted: *2003*

Permission is hereby granted to the University of Alberta Library to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves all other publication and other rights in association with the copyright in the thesis, and except as herein before provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatever without the author's prior written permission.

May 9 2003

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled **Intentions, Conventions, Rules, and Reality** submitted by **Brian Patrick Leahy** in partial fulfillment of the requirements for the degree of **Master of Arts**.

Wes Cooper

Bruce Hunter

Robert Wilson

James Mulvihill

May 8, 2003

Abstract

This thesis examines the role of collective intentionality in creating and preserving social institutions. I present two theories of the construction of social reality: one offered by John Searle in his book, The Construction of Social Reality, the other developed from David Lewis' account of the role of conventions in language. I argue that the only substantial difference between the two theories is that Searle takes collective intentionality to be a primitive phenomenon, while the Lewisian theory offers a reduction. If we are then to choose between the two theories, we may rest our decision on the plausibility of reductionism in this domain. I then argue that the Lewisian does not sufficiently capture the nature of collective intentionality, and conclude that if we are to choose between these two theories, we must side with Searle.

Table of Contents

Introduction.....1

Chapter 1.....5

Chapter 2.....32

Chapter 3.....50

Chapter 4.....75

Bibliography.....107

List of Figures

Figure 1.....33

Figure 2.....33

Figure 3.....35

Introduction

0.1.1. Introduction

We live in a world of molehills, electrons, and gravity. The same world is populated by weddings, parking tickets, and undergraduate degrees. Exhaustive extension of these lists would delineate two fundamentally different sorts of things, and the impetus that led to this thesis was a desire to say clearly wherein that difference lies. For now, I will say this much: members of the former group fit into the causal order unintentionally, which is to say that they have their causal powers independently of what people intend to use them to do. The latter group, on the other hand, may or may not have causal powers independently of what people intend to use those things for. More interestingly, the primary criterion for membership in this group is that an object does have some powers that do depend on the intentions of people in a society to use that object to perform a certain function. Molehills have their causal powers (for example, their ability to keep rain off the heads of their occupants) purely in virtue of their form and the physical laws that govern that form. Money may similarly function causally in virtue of physical laws, such as when coins are attracted to magnets, or independently of our intentions to stratify social classes, but it depends on our intentions for its function as a medium of exchange. Finally, weddings perform no function in virtue of causal laws. Weddings can only give us the privileges and responsibilities they do in a context where it is accepted that a wedding comes with those privileges and responsibilities.

Weddings, parking tickets, undergraduate degrees, and all phenomena that derive powers from the intentions of their wielders are part of what we call the social

world. But the social world is broader than the world of phenomena with such derivative powers. There are all sorts of social phenomena, such as ants relocating from a damaged colony, mother elephants spraying the hot backs of their children with river water, and even a pair of people going for a walk together, that need not require any sort of collective intention, at least not of the sort that generates the institutions mentioned at the beginning of this paragraph. Thus we might pretheoretically distinguish 'social reality' from 'social phenomena' by suggesting that the former are the subset of the latter that depend on our intentions (what we collectively accept) for their existence. So it should be clear that an adequate account of social reality will include a story about the nature of intentions and how we can use them to create social reality.

This thesis is, therefore, an investigation of the role of collective acceptance in generating social reality. First of all, what is collective acceptance? What does it have to do with our intentional states, and how is it related to individual acceptances and intentions? Must we explicitly (intentionally) accept anything at all in order to construct the social world we live in? One might hope not, since such explicit actions seem far too infrequent for the generation of our massive social world. I will carefully outline the role of intentionality in generating social reality. But more importantly, I want to examine the relationship between the intentions of collectives and the intentions of individuals. Producing a social world, as I've already pointed out, depends in some central way on the intentions of the would-be users of that social world. But we might wonder if the intentions that generate the social world differ in type or merely in content from our other intentions as individuals. Does my intention to start a new club with the other philosophy students differ in any way (except its conditions of satisfaction) from

my intention to go to sleep before eleven PM? Several theorists have argued that our social reality demands a form of collective intentionality that cannot be reduced to the intentions of the individuals in the collective. Collective intentions have a fundamentally different structure from individual intentions; they are a different species of intentional state. I will shortly present two views on the role of collective intentionality in generating and maintaining social reality, and ultimately decide that the reductive account does not show as much promise for success as its non-reductive counterpart.

A reductive account of collective intentionality expands the explanatory power of the received view of rationality. It is for this reason that I was first attracted to the notion of collective intentionality. In non-reductive accounts of collective intentionality I find subtle challenges to the received view, and since I am strongly committed to that theory I find myself compelled to marshal arguments against such non-reductive accounts. I have not been successful so far, and unless new information comes to light, I am forced to alter my commitments.

This thesis is a comparison of two possible views of how social reality gets created. One, John Searle's, suggests that collective intentionality is a primitive, irreducible phenomenon; the other, based on David Lewis' account of conventions in language, does not. I compare the two theories on several dimensions, and find that they differ seriously only in this respect. Since this is the only substantive difference between the two, I take it as a criterion for choice between them. I first argue that Searle's arguments for the irreducibility of collective intentionality do not seem to work against the Lewisian account, and leave him with the burden of proof. Furthermore,

there are good reasons to think that the Lewisian account does leave us with a satisfactory reduction of collective intentionality. But a further look at the literature surrounding collective intentionality turns up a counter-example to Lewis' account that shows its insufficiency, and thus the issue is decided. If our choice is between Lewis' and Searle's theories, we must side with Searle.

Chapter 1

1.1.1. The Question About Social Reality

In The Construction of Social Reality[1] John Searle asks, "How do we construct an objective social reality?" (xii). He is interested in the nature of a particular set of objective facts and how they fit into the biologically natural order: what makes it true that this is a five dollar bill (rather than a plain piece of blue paper), that my parents are married (rather than co-habiting), or that I've received an invitation to the department cocktail party (rather than a meaningless card)? According to Searle, the difference between blue paper and five dollar bills depends on what we are willing as a society to collectively accept. This sort of blue paper counts as a five dollar bill in our society because we collectively treat it like a five dollar bill. But why is it that some facts can by made true by such agreement while others can't? No amount of collective agreement will make a mole hill into a mountain. Searle's project in this book is characterizing the class of entities that are and can be constituted by such agreement. In this chapter I aim to familiarize the reader with Searle's characterization. In the remainder of this section I'll offer a brief rundown of Searle's view; following that I'll go into greater detail.

Quickly put, Searle holds that social reality is produced by the collective acceptance of a series of interlocking and iterated constitutive rules of the form, 'x counts as y'. Thus certain extant objects are awarded a special status, and the bearers of those objects are awarded special powers. They are enabled to do certain things under the auspices of *legitimacy*. For example, a certain sort of paper with a certain history counts as a drivers' license in our society. People who hold such licenses are entitled to

[1] I'll refer to this by 'CSR'.

drive. A drivers' license is a piece of social reality, and it bears certain powers. Those powers are determined by what we as a society collectively accept.

1.2.1. Six Features of Social Reality

We can generally find six characteristics of social construction. First, when by collective intentionality we assign a function to something that already otherwise exists, we provide it with powers that it does not have in virtue of its physical form (such as the power of a certain sort of paper (money) to legitimize the exchange of goods). Second, the assignment of this new status function can be represented as having the form 'x counts as y'. The 'counts as' is crucial since we are assigning powers that the object does not have in virtue of its physical form, and requires our collective acceptance in order to get those powers. Third, this whole process might take place without us actually being aware of it. Indeed, this is usually the case. We need only *in fact* treat the x-term object as though it has y-term powers; we need not realize that our actions have that description.

Fourth, assignment of function through collective acceptance gives the 'x counts as y' formula a normative status: it is a constitutive rule. This is proved by the fact that the existence of such formulas introduces the possibility of abuses such as counterfeit money (appears to meet the x-term requirements, but in fact does not) or fugitives (ought to have the y-term status 'is an imprisoned convict' but in fact does not). This is typical of institutional facts: the existence of the rule introduces the possibility of abusing the rule. A surgeon is not only a person with certain skills, but also someone who has a special license; it is because of this license that people are able to masquerade as physicians though they may lack not only a valid license but also the required skills.

Fifth, though it is a matter of rule and not convention that a certain object has certain powers, it is a matter of convention which object has those powers. We call this blue paper that is issued by the government a five dollar bill, but we could just as well call that pink paper issued by the mafia five dollars. Which paper we actually choose, or what we treat as the x-term in any constitutive rule, is up to us within a certain range.

Finally, there is a special relationship between an imposed status-function and language. The labels we give the y-term ('money', 'attorney', 'parking ticket') are in part constitutive of the fact created. We define the term 'money' through the constitutive rule that creates money. The concepts we express in the constitutive rule are the ones we create with the constitutive rule.

I'll now move on to carefully filling out the more important details of this account. I'll begin by rehearsing several distinctions Searle draws on in the course of his book.

1.2.2. The Epistemic and the Ontological Senses of the Objective/Subjective Distinction

Searle thinks that it is clear that there is an objective fact of the matter about whether this is a five dollar bill. But there is also an element of relativity in this fact. To eliminate this apparent contradiction Searle appeals to a distinction between the ontological and epistemic senses of the objectivity/subjectivity distinction. A fact is epistemically objective if it is determined independently of the emotive responses, personal prejudices, etc., of any particular individual. The facts of the matter about epistemically subjective claims are determined by such considerations. An example of the former sort is, "The *Mona Lisa* was painted by DaVinci." An example of the latter sort is, "DaVinci was a better painter than Reubens." A fact is ontologically subjective

if it is in part constituted by subjective mental experiences, and is otherwise objective. Examples of the former are facts about pains, beliefs, and desires, where the state cannot exist if the mind it inhabits does not. On the other hand, a fact is ontologically objective if its truth is mind-independent. If all minds disappeared, there could still be ontologically objective facts such as that the Pacific Ocean is larger than the Indian Ocean.

This is a pivotal distinction in understanding Searle's account of social reality. He wants to "investigat[e] the status of the facts in virtue of which our statements about social reality are true" (CSR 199). Sentences about social reality are true or false in virtue of facts, and those facts are epistemically objective (not determined by the mental states of any particular individual) and ontologically subjective (cannot exist if minds do not).

1.2.3. The Structure of Facts in the Social World

Searle also offers us a partial taxonomy of facts. Epistemically objective facts can be broken up into two sorts: the ontologically objective and the ontologically subjective. The former are called **brute facts** and the latter are called **mental facts**. Mental facts can either be **intentional** or **non-intentional** where the former have propositional content and the latter do not. Intentional mental facts are then divided up into the **singular** and the **collective**. Singular intentional facts imply the existence of the mental states of a single individual, such as Pat's desire for a new pair of pants. Collective intentional facts imply mental states in at least two beings, such as the two musk oxen walking to the river together.

There is a special subclass of collective intentional facts that involve the assignment of function. Assignment of function is a complex notion that I will not develop carefully here. Briefly, the idea is that functional roles are assigned to objects according to our own interests. Thus, resistors are for slowing the speed of electrons, screwdrivers are for screwing in screws, and hearts are for pumping blood. There are important differences in all three cases, but one similarity is that if we were interested in different effects of the object that plays the functional role, we would say it had a different function. For example, if we thought that oscillations from expansion to contraction appeased the Gods, we might say that the function of the heart was to protect us from their wrath by constantly expanding and contracting. Thus, says Searle, functions cannot be not ontologically objective, because an object only has a determinate function relative to the interests of a mind observing that object and its interactions with its environment[2].

There are two broad types of assigned function: **agentive** and **nonagentive**. Nonagentive functions are functions assigned to processes at work in nature, such as the function of the heart as pumping blood. Agentive functions are the functions assigned to objects that are invented by agents for a particular purpose, such as a screwdriver. Hearts might pump blood even if there were no agents trying to make them pump blood. But without agents to use screwdrivers for screwing things in, screwdrivers do not function to screw things in.

There are two types of agentive functions. Objects that can perform their function purely in virtue of their physical features as wielded by agents have **causal**

[2] The point is, of course, contentious; see, e.g., Dretske's "If You Can't Make One, You Don't Know How it Works." I need not take a stance on the issue here, since none of the assigned functions I will be talking about could be realistically construed as natural functions.

agentive functions. But there is a very interesting class of objects that are assigned functions that cannot be performed merely in virtue of the physical features of that object; something more is needed. This is where we assign **status functions**. Screwdrivers work by exerting perpendicular force on a rotating surface. But money does not perform its function (a medium of exchange) purely in virtue of any such physical process. There must be widespread agreement that objects with a certain physical description will play that role. But it does not particularly matter what physical description we decide on: only that we decide. Though we have chosen a certain blue paper to act as a five dollar bill, we might collectively change our minds tomorrow and start accepting instead pink polished stones. This is importantly not the case for objects with causal agentive functions. No matter how firmly we make up our minds about it, foam screwdrivers will not do the trick. Status functions can be generally described as having the form 'x counts as y' where the x-term is to be replaced with the set of properties of the objects we wish to impose a status function on (paper, blue, picture of Laurier, etc.) and the y-term is to be replaced with the new status-function (five dollar bill). When a sufficient population accepts that 'x counts as y', it simply is the case that x is y.

Finally, some of our status functions are **linguistic** and some are **nonlinguistic**. Linguistic functions are imposed on objects (sound patterns, marks on paper, etc.) that will bear semantic content. Such objects gain the capacity to represent things that they are not. Nonlinguistic institutional objects have no such representative capacity. Importantly, however, both sorts of status functions involve linguistic representations as constitutive elements, and those linguistic representations have a particular logical

structure. A large part of Searle's project is saying exactly what he means by this claim, and filling out the details of what that logical structure is.

1.2.4. Physical and Conventional Powers

Searle also distinguishes between physical and conventional powers. **Physical powers** are powers to interact with a purely physical world in certain well-structured ways. These are the powers of hearts, resistors, and screwdrivers. Technological advancement is essentially our growing ability to harness the physical powers of objects as a product of our growing understanding of the natural world (94). But while technological advancement helps us employ physical powers to greater effect, it has no effect on our abilities to harness conventional powers. **Conventional powers** are powers to interact with a social world in certain (perhaps less well-structured) ways. These are the powers of words and weddings. We can't make something a quark just by treating it like a quark, but we can make something a five dollar bill by deciding what (physical) properties a five dollar bill ought to have, and then treating anything that manifests those properties like a five dollar bill. So the question becomes, what are the "forms and limits" (95) of the powers we can impose on things by acting like they have those powers?

This question is essentially equivalent to "What sorts of power can we create just by collective agreement?" (96). The general answer is that we can use the 'x counts as y' mechanism to create all forms of power where sufficiently widespread belief that an object has a certain power is constitutive of the object having the power. Even this general answer is not as empty as it might seem: first, it does not restrict the subject matter of social reality, which explains why the social world is so diverse. Second, it

does not require that the people who are assigning conventional powers realize that they are assigning conventional powers. They need only believe that the object fulfilling the x-term *has* those powers.

Still, we should not be satisfied with this answer to our question. We might interpret the question as demanding a taxonomy of status-functions. On his way to offering such a taxonomy Searle points out that all status-functions have the effect of giving some power to some person. As he writes, "the content of the collective intentionality in the imposition of the status-function will generally be that some human subject, singular or plural, has some power, positive or negative, conditional or categorical" (98). This may be direct, as in the case where we decide to call some immigrants 'citizens' or indirect as in the case of the five dollar bill, since what matters about five dollar bills is what we as agents can use them to do. He goes on to argue that there seem to be four ways we can empower agents via status-functions: we can give them symbolic, deontic, honorific, or procedural powers.

1.2.5. Four Types of Status-Function

Symbolic powers are held by the linguistic status-functions. The string of characters 'Det regnar' counts as a bit of syntax to Swedish speakers. On this particular bit of syntax is also imposed a semantic status-function: it means, 'It's raining.' The existence of symbolic status-functions enables agents to communicate through representations. **Deontic** powers are particular rights and responsibilities imposed on agents as status-functions. Let's consider a few cases of the social situations that involve the imposition of deontic powers:

1. John won the lottery.

2. Suni was elected Student's Union President.
3. Elmer was convicted of armed robbery.
4. Jill lost her license for impaired driving.

In these cases we see the two sorts of deontic powers at play. In the first two cases people are enabled to act in certain ways: John is entitled to payments from the lottery corporation and to buy things he could not previously afford. Suni is entitled to speak on the students' behalf at meetings of the board of governors. The next two cases involve restrictions on the agency of the people mentioned. Elmer is obligated to reside in a penitentiary for some period of time; Jill is no longer allowed to operate a motor vehicle.

An **honorific** status-function provides a status that is valued (or disvalued) for its own sake. By definition they come with no further powers. Examples of honorific status-functions are being named 'champion' of a recreational basketball tournament, being named 'most likely to succeed' in your high school class, or being chastised by your grandmother for speaking out of turn. Finally, **procedural** status-functions are steps along the way to some other (honorific or deontic) status-function. Putting the ball in the net (scoring two points) is a procedural step towards winning a basketball game, but outside the context of the game the points are worthless. Scoring points is only valuable insofar as it contributes to winning a game. Being named a candidate for a position may give you rights and responsibilities, but we are only interested in having those rights and responsibilities insofar as they are a step towards winning the position.

But although we can recognize these four varieties of status functions, all four can be assimilated under the 'deontic' category. Procedural status functions, it turns

out, are just conditional status functions. Scoring two points in a basketball game is a procedural step to winning the game, but we can also cast it thus: scoring two points has the deontic or honorific power of making you the game winner on the condition that the other team doesn't score more points than you. Honorific status functions are a limiting case of the deontic: the power conferred by an honorific status function is just status, which is valued for its own sake, rather than for the sake of some further power attached to the status. Finally, symbolic status functions are like any other status function: they enable an agent to do something she could not do if that status function did not exist (in this case, communicate something). So symbolic status functions are deontic status functions.

With the internal structure of status functions somewhat examined, we can go on to ask exactly how it is that collective agreement manages to assign these various powers to objects. First, Searle notes that all power is power to do something or to prevent something from happening. It follows that all attributions of powers must have propositional content. That content will fit into the schema (S does A). In the case of status-functions the 'S' will be replaced by an individual or group of agents, and the 'A' will be replaced by a description of an action. This description may be positive or negative: being named valedictorian may come with the power (you address the graduating class), and being named class clown may come with the power (you refrain from sitting near other troublesome students). Furthermore, since the sorts of powers determined by status-functions are powers to do things legitimately, the descriptions of actions given in the schema should be prefaced by a qualification of legitimacy.

Anyone who can get past security can address the graduating class; only the valedictorian can do so legitimately.

Furthermore, since (as we noted above) conventional powers are the forms of power where sufficiently widespread belief that an object has a certain power is constitutive of the object having the power, the power is constituted by a collective intention of the form, 'We accept (S has the power (S does A))' (104). This is equivalent to the conjunction of 'We accept that x counts as y' and 'We accept that the wielder of y has the power to do A'. Combining the two, we get, 'We accept (S, as the bearer of x, which counts as y, has the power to do A).'

We might then ask how it is that mere acceptance that a person has some power can turn into that person having that legitimate power. When we undertake certain actions in an appropriate context, we commit and entitle ourselves in certain ways to certain things. These commitments and entitlements are determined by what is collectively accepted by a society or substantial portion of a society. For example, uttering certain words in front of a priest in a certain formal setting counts as getting married, and we allow married couples the conventional power of a sexual monopoly. The commitments and entitlements of a particular social action in a particular context will be specified in the culture for every such action-context pair. Collective intentionality allows us to bestow conventional powers on an event because acceptance that an event has certain conventional powers constitutes that event as having those causal powers. When S, the bearer of x, which counts as y, does A (the action she is entitled to according to our collective intentions), her action is legitimate. If the person

who does A is not so entitled by our collective intentions, the action is illegitimate. The legitimacy of an action is a constituted by us calling that action legitimate.

1.3.1. The Role of Language

Searle notes that it would be impossible to construct this sort of epistemically objective/ontologically subjective fact in a population without language. When we decide to hold that x counts as y, we are doing something that is essentially linguistic, because the y-term cannot exist outside of a linguistic framework. A non-linguistic being could not think 'that's a five dollar bill' or 'that's a touchdown' because five dollar bills and touchdowns can only exist in linguistic contexts. To see this, ask what it would take for a non-linguistic being, Angus, to recognize that Weston has just scored a touchdown. Clearly Angus could watch Weston catch the ball, evade the defence, and cross the line. But perceiving these events *as a touchdown* requires language-it requires knowing the rules of football, which is a propositional form of knowledge. It requires familiarity with the system of scoring and keeping track of points, which is necessarily a linguistic system. We can't go out into the world and find *points*. Points (five dollar bills, parking tickets, weddings, etc.) only exist insofar as we *call* something a point (a five dollar bill, a parking ticket, a wedding, etc.). Calling something (x) something else (y) is something you can only do if you speak a language.

1.3.2. The Role of Collective Intentionality

Searle is then entitled to conclude that agreement that an event has certain conventional powers is equivalent to that event having those conventional powers. But central to this argument is the notion of collective intentionality. Without collective intentions we can't give anything or anyone conventional powers. But what are these

collective intentions, and where do they come from? Though I'll develop Searle's account of collective intentionality more thoroughly in chapter 3, a brief answer to these questions is useful for filling out the major theoretical points at work in The Construction of Social Reality.

Sometimes agents act alone, and sometimes they act in groups. Thus we say, "I am stuffing the turkey" while "We are preparing for our Thanksgiving dinner" or "I am writing a paper" while "We (my supervisor and I) are trying to get me through my MA degree." All examples of people acting in groups involve co-operation to some common end. This is true even when people are co-operating in order to conflict: for example, in a political debate. The ultimate end of each politician is exclusive of the end of the others, since all of them want to win the election but only one can. But they all want to have a debate, and in order to have a debate they must co-operate. At the very least they must all agree to be in the same place at the same time.

Searle argues that this co-operation involves collective intentionality. Agents, when they act alone, intend alone: they have "I-intentionality." But when agents act together, they intend together: they have "we-intentionality," or collective intentionality. This collective intentionality is not reducible to individual intentionality, not even individual intentionality augmented by anything else such as beliefs or mutually held beliefs. Collective intentionality is biologically primitive. But importantly, Searle does not want to be committed to any sort of 'collective mind' that somehow hovers over and above the minds of individuals. But he does think that individuals' intentions can refer to the actions of collectives. So, I might have the intention in my head, "We intend to prepare our Thanksgiving dinner" and it is this

intention combined with your type-identical intention that produce the collective action of us preparing for our Thanksgiving dinner.

For Searle, collective intentions are the product of a special sort of co-operation between agents. For example, when we say that we intend to prepare our Thanksgiving Dinner, each of us has an intentional state whose content is partly specified by that proposition. The rest of the content is specified by a proposition that determines our particular efforts to that common goal in a 'by means of' relationship. So, the complete content of my intentional state might be, "we intend to prepare our Thanksgiving Dinner by means of me individually plucking the turkey." Everyone involved in the collective action will have an intention with the same we-content and an individual-content that determines their individual actions towards that common goal. The individual-content is derived from the we-content in a way that we will make clear later, but for now we might mention that constructing social reality requires that we have intentional states with this collective by-means-of structure. So, when we say that "We collectively accept that this sort of paper counts as a five dollar bill," we must all have an intentional state of the form, "We collectively accept that this sort of paper counts as a five dollar bill by means of individually acting as though this were a five dollar bill". Any theory of intentional content that tries to reduce the 'we'-content to some combination of I-content and something else individualistically stated (perhaps beliefs about the actions of the others in the group) necessarily fails. We'll see Searle's arguments for this in the third chapter.

1.3.3. Rules and Social Reality

Constructing social reality is a rule-governed activity. Specifically, status functions are assigned by constitutive rules. Constitutive rules, unlike regulative rules, bring into existence things that can only be brought into existence by collective agreement. For example, the rules of basketball allow basketball to exist. Without such rules, basketball simply does not exist. Regulative rules, on the other hand, govern some already existing phenomenon. Constitutive rules always have the discernable form 'x counts as y'. Collectives who wish to construct or continue the existence of something (for example, basketball) must collectively accept a set of rules that determines what basketball is.

Everything constituted by constitutive rules is a social institution, and what goes for basketball goes for all social institutions. When we collectively agree that a certain event confers some rights and obligations on a person, each individual has an intentional state of the form: "we intend to accept that x counts as y by means of individually acting as though x is y" where the x-term names a physical description of an event and the y-term denotes the status function. When acceptance of the constitutive rule is suitably widespread, it simply is the case that x is y. Of course, no one might actually be aware of any intentions of this form; Searle resolves this problem by appealing to his thesis of the Background, which I'll explain shortly. The value of this formulation lies in its ability to identify structure in our very complex world of social institutions. Our various social institutions are so interdependent, in fact, that it is difficult to find a non-artificial example of a social institution that does not somehow presuppose the existence of other social institutions. For example, part of the context

that allows saying, "I do" to count as getting married in our culture is that those words are uttered in the presence of an appropriate official. But being an official involves bearing a status function. Officials are socially constructed.

The point is that the structures that yield social institutions are interlocking and iterated. The example above of marriage's dependence on the presence of an appropriate official demonstrates how social institutions interlock. Furthermore, the x-term in a marriage (saying "I do") is a speech act. Speech acts themselves are linguistic status functions, which shows that status functions can be imposed on status functions. Status functions can be iterated. The formula "x counts as y" allows us to easily distinguish the various social institutions at play in any social event.

It is important to note, however, that for Searle brute facts are logically prior to institutional facts. This means that no matter how complex our networks of status functions, the web still has to be tied to the physical world somewhere. Every status function *a* is imposed on an object *b*, which either already had a status function or did not. If it did not, then we see that *a* is tied to the physical world. If it did, we must look at the object *c* on which *b* was imposed, which either already had a status function or did not. If *c* is not a status function, then *b* and therefore *a* are both tied to the real world. We can continue through this genealogy of impositions of functions for any status function until we come to an object that is not itself a status function. This must always be the case since status functions must always be imposed on something; there must be something in the role of the x-term.

1.4.1. A Review

Let me restate clearly Searle's main points. First, we should remember that social reality is constructed through a system of constitutive rules. Thus, to say that this piece of paper is a five-dollar bill is to say that we accept that this sort of paper counts as a five-dollar bill. To say that this is a wedding is to say that this set of actions in this context counts as a wedding. But 'counting as' is a function of what is collectively accepted: it means little more that 'is treated like' or 'is taken to be'. Thus Searle notes the 'self referentiality' of many social concepts. By this he means that the claim 'x is money' implies an indefinitely long disjunction to the effect that 'x is treated like money or x is believed to be money or x plays the role of money.' This implication shows that, when it comes to socially constructed objects, you simply can't fool all of the people all of the time (32). If no one treats a certain type of object like money, then it simply is not money; if everyone (or even if enough people) treats a certain sort of object like money, then it is money-ie, a repository of value, a medium of exchange, etc.

Here we see the role of collective intentionality: having a certain status-function is just to be the thing that people accept as having that status function. Collective acceptance that this object *x* should be have this function *y* works because a big enough group of people has chosen to limit their own agency in certain ways. Accepting that this paper counts as money involves accepting that if I don't have any of this paper, I won't be able to do certain things (go to the movies, buy dinner, pay my rent). Accepting that there should be a Prime Minister involves restricting my agency in ways compatible with having a Prime Minister: I must consent to be governed, to pay my taxes, to follow the laws made by the government of which the Prime Minister is the

head. Accepting these limits on my agency is the same as accepting that there will be a Prime Minister, since if we do not jointly accept the limits imposed on our agency by a particular conception of 'Prime Minister' then the Prime Minister will not be able to act as Prime Minister, and hence will not be a Prime Minister. This is why when societies re-organize and start accepting different sorts of limitations on their agency we experience social change-like revolutions.

1.5.1. The Non-Intentionality Problem

It should be objected that there is something strongly counter-intuitive about Searle's account: it wants to call our acceptance of certain objects as having certain conventional powers 'intentional', and suggests that we can only create social reality via very specific intentional actions. But it seems that, while we do have an (epistemically) objective social reality, we rarely if ever perform the intentional acts demanded by the model (at least in any interesting sense of 'intentional'). No one *intends* to accept that this sort of paper is money; we simply grow up in a situation where certain sorts of paper *are* money, and we go along with that. Searle is prepared for this objection. To answer it he relies on his famous thesis of the Background. In the next section of this chapter I'll explain the nature of the background; in the section thereafter I'll describe the subtle way in which Searle uses the thesis of the background to resolve this problem. I'll call the problem 'The Non-Intentionality Problem'. It is important to note at this point that the non-intentionality problem is a problem for how social reality is perpetuated, not how it is created. Presumably a great deal (if not all) of our creations of institutions require intentional acts: if we want to create a club, a new

game, or a new sort of wedding, we have to intentionally 'write' the constitutive rules for that new institution.

1.5.2. The Thesis of the Background

The Background, writes Searle, is a set of non- and pre-intentional capacities that are required to allow our full-blown intentional states to function (CSR 129). In Searle's earlier work the non- and pre-intentional states that constitute the Background were distinct. The former were called the Background, and the latter were the Network. The Network is a set of intentional states that provide context for a given intentional state. A given intentional state can only make sense as an intentional state within a network (context) of other intentional states. For example, my desire for two pounds of firm, ripe tomatoes does not determine conditions of satisfaction if I have no other beliefs about things like what tomatoes are, what counts as a ripe tomato, how much two pounds is, and so on. But even this Network is not enough to determine the conditions of satisfaction for the given desire: we also need the Background. Intentional states, even in the context of a Network, are not self-interpreting. There is a radical underdetermination of the content of any intentional state by the descriptions we give of those contents. Even if it is given that I want two pounds of firm, ripe tomatoes, nothing in the literal meaning of the description of the content of that desire blocks the interpretation that I want those tomatoes to fall from heaven and land on my petunias, or that I want the tomatoes to come and take me away to a happier place. Adding more Network intentional states does not solve this problem, since those states are equally open to interpretation. It is because of our Background of non-intentional states that we find such interpretations unlikely. Our Background capacities are bits of know-how

that tell us what sort of a desire it should be when someone tells us they want two pounds of firm, ripe tomatoes.

Since The Rediscovery of the Mind, however, Searle has modified this understanding of the Background. Now it is taken that the Network is part of the Background. Furthermore, he is now careful to avoid talk of the Background as a 'set' of states at all, as it is a mistake to think that somewhere in the brain there is a sort of library of representations where we can go and check out the mental states we need right now and make them conscious. Rather, we should understand the Background as a mechanism for generating sets of skills and conscious mental states when we need them. The Network, rather than being a group of intentional states that helps determine the content of our occurrent intentional states, is instead the ability of the Background to generate lots of other conscious thoughts. So, to believe that dirty oil is black when one is asleep is to have a brain structure or process in place that is able to produce the appropriate intentional state when it is called for[3]. Hence the capacities Searle is interested in here are causal structures of our neurophysiological wiring, which are so wired as a product of our learning to navigate the world (more on this later). But since we don't know enough about neurophysiology at this point, we describe these causal structures at a higher level. For example, the ability to speak English is in all likelihood a neurophysiological structure. But since we don't yet know the details of the neurophysiological realization of that ability there is no serious problem in identifying it

[3] I should be careful to point out that this claim is not meant to commit Searle to the identity theory. Searle's view is that conscious states are emergent properties of brain states. An emergent property is one that cannot be calculated from complete knowledge of the elements that compose an object without an account of the relations between those elements. For example, no matter how much you know about the molecules that compose a bit of water, if you don't know anything about the relations between those molecules then you don't know whether the water is solid, liquid, or gaseous. The state of a bit of matter is an emergent property. On the other hand, some view this emergentism as a form of the identity theory; see Joseph Levine "Recent Work on Consciousness" note 3.

as 'the ability to speak English.' These causal structures are necessary for the function of intentional states in a purely physical sense: they are at play in the physical causation of our intentional states.

To really appreciate the value of the Background Searle recommends that we consider cases like the English verb 'to cut.' The verb functions in many ways. It requires that we do very different things in each of the following commands: 'Cut the grass!' 'Cut the ribbon to open the building!' 'Cut the pie into equal sized pieces!' In each of these sentences 'cut' is being used literally and without any lexical ambiguity; the meaning of the verb is constant. We can see that there is no lexical ambiguity because it makes perfect sense to say, "I bought a new machine that cuts grass, pies, and ribbons at opening ceremonies!" But when we get into lexical ambiguities, such conjunctions no longer make the same sense. It makes no literal sense to say, "I bought a machine that cuts grass, pies, and spending in the third quarter!" (Rediscovery 179). Nonetheless, the verb 'cut' determines very different conditions of satisfaction in each of the lexically unambiguous cases above. There may be an important sense in which I have not followed orders if I perform the ribbon cutting ceremony with a lawn mower, though a lawn mower is an appropriate tool for cutting lawns. But the literal meaning of 'Cut the ribbon to open the building!' does not block the lawn mower interpretation. That interpretation is instead blocked by the Background: non- or pre-intentional states that just are the ability to interpret the context-sensitivities of utterances. The Background is non-intentional know-how, like the things we learn when we learn to ride a bike. When learning to ride a bike we do not first master a set of rules and then apply them; rather, our ability grows organically through practice, trial, and error.

Though there may be explicit principles that govern bike riding, we need not have those principles as the content of any of our intentional states (Rediscovery 185). For example, we need not believe that when the bike lists to the left, I can prevent myself from falling by turning the wheel to the left and shifting my weight forward. Certainly many people are able to rescue themselves by exercising such know-how without having any beliefs to that effect. And what goes for riding bikes and interpreting sentences goes for interpretations generally, including interpretations of perceptual data. Perception is always perceiving as: perceived objects are always categorized under some more or less familiar concept (CSR 133). But such categorization demands the application of Background abilities.

1.5.3. The Background and the Non-Intentionality Problem

This notion of the background gets Searle out of the non-intentionality problem by allowing that our rule following needs to be neither conscious nor unconscious in order to perpetuate social reality. He argues that the Background causes the behaviours that constitute the perpetuation of social institutions, but it follows neither of the traditionally recognized causal routes: the causation is neither mental (ie, intentional) nor physical (billiard-ball)[4]. Yet while our rule following is neither conscious nor unconscious, perpetuating social reality is still a rule governed[5] activity. This is odd; we might wonder how our behaviours can be rule governed if we are not even unconsciously following rules. Searle responds to this query by first pointing out that

[4] Of course, mental causation may be a variety of physical causation

[5] Rule governed behaviour is what we have when we are intentionally following the rules, such as a beginner playing a game of chess. Rule described behaviour is when there is no intentionality that generates compliance with the rules. It is a rule that objects falling near the earth's surface accelerate at a rate of 9.8 m/s squared, but the rule has no role in producing the behaviour. The behaviour is produced by interactions between forces that are regular in ways such that we can call them 'laws.'

our behaviour cannot be merely rule-described, since the rules do have some causal role in producing behaviour. He argues that there is a third type of causal role (in addition to mental and physical causation) that rules can play in modifying our behaviour: Background causation.

Searle argues that understanding this new type of causation demands an inversion in our order of explanation similar to the one instigated in biology by the theory of evolution. Before evolution biologists might have said, "The fish has the shape it does so that it can swim better, that is, so that it can survive." There is a teleological structure to this sort of explanation of survival: survival is a goal at which (adaptive[6]) features of the organism were aimed. Since Darwin and the development of genetics, we have learned that a better explanation is "The fish has the shape it does because of the way its genes interact with the environment, and many fish have those genes because that genes-environment combination (and that shape) allow fish with those genes to survive and reproduce." Searle's argument for Background causation employs a similar inversion. Rather than saying that people are intentionally following rules in order to perpetuate social institutions, we can say that we are neurophysiologically arranged so that our behaviour follows certain patterns (which, by the way, perpetuates the social institutions of the society), and we are so arranged because it fits with the structure of the society into which we were born (CSR 144). The relevant neurophysiological arrangement is the brain mechanism that Searle calls the Background which, as stated above, is part of our neurological wiring, but we become so wired through an organic process of non-rule based learning: the learning is

[6] Brackets are used here because I recognize the anachronism of my use of this term here. Still, the term refers to the right properties, even if they were not so identified before Darwin.

a product of social conditioning. We learn by trial and error how to behave with respect to our social institutions. The case parallels learning to ride a bike: being given a set of axioms will help no one get the bike going. Getting the bike going requires getting on the bike and refining your movements until the bike stops falling over, whether you have a set of axioms or not. Similarly, when we are born into a culture, no one gives us a set of rules that we have to follow in order to perpetuate the social reality of that culture. We are sensitive to our environments, including the social constituents of our environments. This sensitivity allows us to learn how to properly interact with the social world, and properly interacting with the social world (following the constitutive rules) is the perpetuation of the culture's social reality.

There are two important parallels between the case of evolution and the case of the Background. First, both models provide an explanation for why things are they way they are (in the first case, genes in interaction with the environment cause a particular shape; in the second case, neurological organization in interaction with the environment disposes one to particular sorts of behaviour). Second, both models eliminate teleological explanations: survival is no longer a goal, it is a by-product of certain genetic configurations; the perpetuation of social reality is often not a goal but a by-product of our accepted ways of behaving (e.g., using money, issuing parking tickets, etc.). It may require intentionality to explicitly create a new bit of social reality, such as when we found a club, but our perpetuation of that entity is rarely a goal of but rather a by-product of our participation in the club. In one sense, we don't go to club meetings in order to keep the club alive; going to club meetings *is* keeping the club alive.

With such an inversion of the order of explanation, we can see how the constitutive rules of our social institutions have a causal role in producing our behaviour, but the causation is neither mental nor physical (neither intentional nor billiard ball). We do not explicitly learn, practice, and apply the rules intentionally, but rather find some patterns of behaviour reinforced and others punished. Thus as we grow up we learn how to behave, but not by learning to apply rules. We just become familiar with how our various institutions like money and parking tickets work, and we do in fact abide by those rules. So our collective acceptance of the constitutive rules of social institutions need never be explicit, though it is always tacit. Our acceptance is implied by our behaviour.

The workings of this theory might seem clearer in the light of an example. Searle tells the story of the development of paper or 'fiat' money in Europe. The story may or may not be historically accurate; more importantly, it develops all of the important themes in his theory. A long time ago, there was no money as we understand it today. In medieval Europe, for example, gold was often used as a medium of exchange, because that precious metal was taken to have a worth of its own. This form of money is called 'commodity money', because the medium of exchange is itself regarded as valuable. Some time later, banks started storing gold for safekeeping. Bankers would give depositors paper certificates that could later be redeemed for gold. These certificates themselves soon became a medium for exchange, as they were seen as contracts that could be traded for something that was valuable in and of itself. Thus commodity money was replaced by contract money.

Eventually someone realized that there was a way to increase the supply of money: since contracts need rarely if ever actually be redeemed with gold, it is possible to print more contracts than can be redeemed in gold. All that really matters is that there is a 'common currency' with a standard value that is widely accepted. With this acceptance firmly in place, bankers were able to forget about gold altogether and simply print paper money. This final stage of development is called 'fiat money.'

In cases like fiat money, we have a group of people who collectively accept that a certain object has some worth, while that object has no worth independent of that acceptance. But since their acceptance is produced by the background, since it is a form of behaviour that has been reinforced and grown organically rather than taught according to rules, it is a non- or pre-intentional acceptance.

So, we might ask, what is the ultimate picture of the construction of social reality we are left with? Let's keep the example of money. There are three important steps in the preservation of money as a social institution. First, we are born into a social world where people employ money in certain ways. They accept it as a medium for exchange, they hold it as a repository of value, and so on. Second, because this institution is so useful and ubiquitous, we learn rather unwittingly and from an early age how to use money. The learning is unwitting because it does not take place in a principled manner; rather, we acquire a bike-riding style Background know-how. Third, because of that know-how, caused in us by rules in a neither mental nor physical way, we continue to navigate the world in a way that keeps the system going. We keep using and accepting money, and as long as people act as though this type of paper is money, then this type of paper is money.

There are, however, intentional states that go along with perpetuating social reality. Having a social world comes with norms: we think there is something wrong with acting illegitimately. We have to be prepared to offer some degree of censure to those who break the rules of our social world.

1.6.1. Conclusion

In this chapter I have offered no arguments of my own; as a result, I have as yet no conclusions to draw. I take myself, however, to have satisfactorily outlined Searle's account of the position of social facts in the genealogy of facts, and how such facts get created merely be acceptance of various institutions. I have gestured towards the role of collective intentionality in this account. In the next chapter I develop another account of how we get a social world out of David Lewis' account of the origins of language.

2.1.1. Lewis on Co-Ordination

In this chapter I will draw on Lewis' account of linguistic convention to develop an alternative account of the construction of social reality.

On any given day we face a multitude of co-ordination problems. By way of defining 'co-ordination problem,' let me offer a few examples:

1. Will and Nils are planning to spend their day off together. They want to go skiing. They don't care where they go-only that they go to the same place.

2. Will and Nils are planning to spend their day off together. They want to go skiing. Will has a season's pass to Hill A and Nils has a season's pass to Hill B. Each would most prefer to ski with their friend at their own hill than to ski with their friend at their friend's hill. But at the same time, each would prefer skiing with their friend at their friend's hill to skiing at their own hill alone. The almost-worst-case-scenario, only slightly better than staying home alone and doing nothing, is skiing alone at the friend's hill.

3. Will and Nils are looking for a common friend. Their friend is skiing at one of two places, and in order to find him they will have to split up. Neither cares which hill he goes to; only that he goes to the hill that the other does *not* go to.

4. In each of the previous three cases, Will and Nils are trying to figure out where to go. They are having a telephone conversation aimed at solving their problem. They are unexpectedly cut off, so one must call back while the other waits. Neither has a preference for calling or waiting; only that the other does not do the thing he does.

5. There are also co-ordination problems involving many players. Having a unified nation may require that we all (or at least very many of us) recognize the authority of the same leader. It might be the case that no one cares which leader we recognize, only that some one person is endowed with the power to negotiate on our behalf in international relations, etc.

6. The Swiss Family Robinson might land on their island facing a desperate situation. They face extinction unless they find shelter, food, water, and clothing very quickly. Conditions are such that they must split up in order to perform all the tasks in time. No one has any skills or interests that makes them better suited to any particular job.

7. Each member of a group of conspirators must learn a code for secret communication. No one cares which code they learn, but no one has time to learn more than one code.

Hopefully this reveals the great variety of co-ordination problems, as well as their common thread. It is tempting to say that the common thread is that every agent wants to do the same thing as the other agents, under some description of 'same'. This will not do, as Lewis points out: virtually any combination of actions can be described as 'the same'. Consider case number four: are both agents doing the same thing if they both wait for the other to call, or are they doing the same thing if they both wait if and only if the other calls (Lewis 11)?

In order to more clearly say what all co-ordination problems have in common, Lewis resorts to game theory. Every two-player co-ordination problem can be represented as a matrix with the rows representing the possible choices of one player (call her Row) and the columns representing possible choices of the other player (call him Col). In each row/column pair is a set of two numbers. The first number

represents Row's utility for that combination of actions; the second number represents Col's utility. In case number 1 above, we might expect something like this:

		Col	
		Go to Hill A (C1)	Go to Hill B (C2)
Row	Go to Hill A (R1)	1,1	0,0
	Go to Hill B (R2)	0,0	1,1

Figure 1

In this case <R1, C1> and <R2, C2> are the solutions to the problem. These positions are also known as *equilibria.* An equilibrium solution is a set of actions whereby each player could do no better by changing her action, given that all the other actions stay fixed. Given that Col goes to Hill A, Row cannot improve her outcome by going to Hill B; given that Row goes to Hill B, Col cannot improve his outcome by going to Hill A. Indeed, Lewis suggests that what matters to prospective co-ordinators is not any sense of uniformity of action, but action that is in a particular state of equilibrium.

The characterization of that particular form of equilibrium is necessary since we cannot distinguish co-ordination problems from other sorts of games by reference to equilibrium alone. There are many games with equilibria that are not co-ordination problems, such as the following:

		Col	
		C1	C2
Row	R1	1,1	0,0
	R2	0,0	0,0

Figure 2

This is not a co-ordination problem because it neglects the character of working with someone else necessary for co-ordination problems. In this game each player can come

to a rational decision without considering the other agent, since each player really only has one possible move. Row has a chance of return if she does R1 but not R2, so she should do R1. Similarly for Col. This independence shows that equilibrium alone cannot distinguish co-ordination problems.

To narrow the field, Lewis defines 'co-ordination equilibrium'. Where an equilibrium solution is one where no player could do better by changing his act, a co-ordination equilibrium is one where no player would be better off if any single player changed her act. This notion ensures that co-ordination equilibria are solutions where all players must find a way to work together, which explains the intuitiveness of the suggestion that in co-ordination problems all players must in some sense do the same thing.

Furthermore, all co-ordination problems must have at least two co-ordination equilibria. The equilibrium solution in figure two is a co-ordination equilibrium, yet it is still not a co-ordination problem, and that is because the players of that game need not consider which equilibrium solution the other will choose. So the problem is not only that all players must find the same solution, but that they must find that solution as a possibility amongst alternatives. Will and Nils must either go to Hill A or Hill B; citizens of Nation X must designate Candidate Y instead of Candidate Z as their leader. If the choice is between Candidate Y and no one at all, we only have a co-ordination problem insofar as choosing no one at all is a co-ordination equilibrium. Of course, in cases like this last the notion of co-ordination equilibrium needs some refinement, since not all citizens need to recognize the same candidate as leader in order for the game to

be solved; only a sufficient degree of co-ordination is required. I see no reason to think the need for such refinements pose any serious threat to Lewis' theory.

But not all games with two or more co-ordination equilibria are co-ordination problems. Figure three does not represent a co-ordination problem, because neither agent

		Col	
		C1	C2
Row	R1	1,1	1,1
	R2	0,0	0,0

Figure 3

needs to consider the other. Row will choose R1 no matter what, and Col may make either choice. In cases like this, even if they are aiming at different equilibria, an equilibrium solution will be reached (Lewis 21-22). This possibility eliminates the sense of real co-ordination: it seems that the sort of co-ordination we are looking for is in some sense intentional. Players all have to be aiming at a certain solution in order to be really co-ordinating. Co-ordination should not happen by accident.

Lewis eliminates this possibility by defining 'proper equilibria'. In an equilibrium solution, a player can do no better by unilaterally changing her move. That allows the possibility that there are other equilibria available to her that are just as good as the solution reached. Proper equilibrium excludes this possibility: a proper equilibrium is a solution where a player can only do worse by unilaterally changing his move. So, Lewis argues that there must be at least two proper co-ordination equilibria

in a co-ordination problem. This requirement will force all players to think about which choice the other players are aiming at.

There is one other factor that distinguishes co-ordination problems. Psychologically speaking, what matters most to players in co-ordination problems is that they manage to arrive at a co-ordinated solution. Other interests may be at play in determining which solution various players prefer (witness example 5 above), but if, after an election, voters refuse to acknowledge the elected leader and instead try to acclaim the victory of the candidates for whom they individually voted, we no longer have co-ordination. We have several agents who have not arrived at a co-ordinated solution and, if their individual interests are such that the difference in preference-values for particular solutions outstrips the value of co-ordination, they will not arrive at a co-ordinated solution. If a proposed co-ordinated action has a lower expected utility than some viable individual action, then the individual is rationally bound to reject the collective action and instead act individually[7].

2.1.2. Solving Co-Ordination Problems

There are several ways to solve co-ordination problems. Perhaps the most common way is through promising. Consider the first example, where Will and Nils are planning to spend their day off skiing together. They don't care where they go-only that they go to the same place. This sort of co-ordination problem is usually solved when one player says, "Lets go to Hill A" and the other agrees. The possibility of such a solution does not make the problem any less of a co-ordination problem: there are still at least two proper co-ordination equilibria. The only difference is that now each agent has a rather conclusive reason to adopt a particular strategy.

[7] Though, of course, her moral or other commitments may bind her otherwise.

Co-ordination problems are also solved in more subtle ways. Consider the variation of the third example, where Will and Nils are looking for a common friend who is skiing at one of two places, and in order to find him they have to split up, but the phones are out so they cannot communicate. Perhaps they have been in this situation before, and Will went to Hill A and Nils went to Hill B. If both players reason that what worked last time will work this time, then the problem will be solved.

Or, consider the Swiss Family Robinson. They face extinction unless they find shelter, food, water, and clothing very quickly, and conditions are such that they must split up in order to perform all the tasks in time. There may be a particularly salient solution-perhaps each member of the family has a skill that suits them to a particular job. They may fall to an appropriate solution without so much as a word.

Now suppose that each of these last two situations were repeated several times, and each time the same solution was reached by the same reasoning. It might eventually become the case that all players stop thinking about why they choose the option they do (because it happened last time, because of their particular skill) and simply do the appropriate action simply because that is the action that they are supposed to do in that situation. We might say that in this case the players have developed a convention. Conventions allow us to choose between competing proper co-ordination equilibria in co-ordination problems. They are tie breakers.

Lewis examines several examples of conventions like the two discussed above. He arrives at this formal definition: "a regularity *R* in the behaviour of members of a population *P* when they are agents in a recurring situation *S* is a *convention* if and only

if it is true that, and it is common knowledge in P that, in almost any instance of *S* among members of *P*:

1. almost everyone does *R*;
2. almost everyone expects everyone else to conform to *R*;
3. almost everyone has approximately the same preferences regarding all possible combinations of actions;
4. almost everyone prefers that any one more conform to *R*, on condition that almost everyone conform to *R*;
5. almost everyone would prefer that any one more conform to *R'*, on condition that almost everyone conform to *R'*,

where *R'* is some possible regularity in the behaviours of members of *P* in *S*, such that almost no one in almost any instance of *S* among members of *P* could conform to both *R'* and *R*" (Lewis 78, italics in original).

Conditions 3, 4, and 5 ensure that we have a co-ordination problem. That it is common knowledge that conditions 1 and 2 hold ensures that we have a conventional solution to our co-ordination problem. It is common knowledge that x in a population P if and only if there is a state of affairs A such that:

A. Everyone in P has a reason to think that A is true;
B. A indicates to every individual in P that all members of P have a reason to think that A is true;
C. A indicates to every individual in P that x.

State of affairs A is then called the basis for common knowledge. It enables members of P to form beliefs of arbitrarily high order about the beliefs of other members of P:

given some reasonable assumptions about induction and rationality, I can conclude that you know that I know that you know…that x.

So, since it is common knowledge that conditions 1 and 2 hold, there must be some basis A that everyone has reason to think is true and to think that everyone else thinks A is true, and that indicates to everyone first that almost everyone does R (i.e., conforms to the convention) and second that almost everyone expects everyone else to do R. This is exactly what we want of our conventions. The basis A tells each player what they have to do in order to reach a co-ordinated solution and gives her a reason to think that everyone else will play their role.

2.2.1. Conventions and Signalling Problems

Having rigorously defined the sense of convention he is interested in, Lewis proceeds to attack one of the fundamental problems in the philosophy of language. Ultimately he wants to show that language is conventional as it trivially seems to be, despite the sceptical criticisms launched by Quine and others. He begins by showing how using conventions to solve co-ordination problems is in some cases exactly what we need in order to communicate. For example, it is convention amongst ships' officers that you should hoist a white flag to port and a blue flag to starboard if you're on a speed trial; white to port and red to starboard if you're loading explosives, and so on (126). These regularities in behaviour are conventions because it is common knowledge that (1) almost everyone conforms to the convention, (2) almost everyone expects everyone else to conform to the convention, (3) the psychological condition for being a co-ordination problem is met: using the flag patterns others use is what really matters to most ships' officers, (4) almost everyone prefers to conform, on the condition

that almost everyone else conforms, since people don't want to interfere with speed trials (get run into) and they don't want to have their speed trials messed up, and they don't want to cause others to blow up or to get blown up themselves, and (5) there is at least one other proper co-ordination equilibrium R' that almost everyone would prefer to conform to, if almost everyone else were to conform to R'. If there weren't some convention about which flags to hoist in which circumstances, it would be difficult if not impossible to communicate your circumstances to other ships by using flags.

Lewis calls the sort of co-ordination problem we face when we want to communicate some details of our situation to someone else a 'signalling problem'. We solve signalling problems by developing signalling conventions to use signalling systems. In the case above, the signalling system is the set of relations between situation-details and flag combinations, and the signalling convention is that we use that particular signalling system. Lewis defines 'signalling problem' carefully. We have a signalling problem when there is one person, call her the communicator, who is in a position to discern which of several possible states of affairs holds, and another person or group, the audience, who is not in such a position but wishes to know which state of affairs holds. The audience wants to use the information about which state of affairs holds to determine his actions. There is a function that maps each possible state of affairs onto a response from the audience, and both the communicator and the audience want the appropriate response to be given for the state that actually obtains. The communicator also has several action choices, called signals. There is a functional relation from possible states of affairs to signals. The audience is in a good position to

see which action the communicator chooses, and also knows the function from possible states of affairs to signals.

From that information, the audience is able to calculate which state of affairs actually obtains, and from his knowledge of the function that maps possible states of affairs onto audience responses he is able to choose the appropriate action. The function from possible states of affairs to signals by the communicator is called the communicator's contingency plan; the function (derivable from the information above) from signals to acts by the audience is called the audience's contingency plan. If we assume that both communicator and audience must choose their contingency plans from a set of alternatives (this is reasonable-there is no reason to think that any state-symbol or symbol-response pairing is not arbitrary) and then acting according to their chosen plans, we have a co-ordination problem. Both players want to choose a signalling system that is appropriately related to that of the other: they both want to choose a function that relates the same symbols to the same states of affairs. A signalling convention is a convention among a population P that tells them which contingency plan to adopt in such a situation. By following the convention, all players can be confident that all players have chosen functions that relate the same symbols to the same states of affairs.

2.2.2. Signalling Systems and Language

There are good symbols and bad symbols. Good ones are easy to observe, bad ones are not. Good symbols also usually lack value beyond their value as a symbol, so there is no problem with using them as symbols. Two examples of symbols meeting both of these conditions are marks on smooth surfaces that are visible to the audience

and strings of vocalizations within earshot of the audience. Speaking and writing, in other words, are excellent media for signalling systems. Of course, signalling systems of the sort we have here are severely limited when compared with natural languages; people who only speak such a language could only say things that are readily observable, or are associated with some responsive action, or even both. They could also only generate a finite number of sentences, since we have no method for generating new symbols for new phenomena. Lewis expands his account of conventional meaning so that it more resembles natural language, but we need not rehearse those arguments here. Even Lewis' most primitive signalling systems are what Searle calls linguistic status-functions, and so we can derive from Lewis' account of signalling systems an account of status-functions generally.

2.2.3. Conventions, Signalling Systems, and Social Reality

We have seen that signalling systems, a variety of Searlian linguistic status function, are conventions that allow us to resolve signalling problems, which are a species of co-ordination problem. Mutatis mutandis, Lewis' arguments could be used to establish that all status functions are conventions that allow us to resolve some species of co-ordination problem. Money, for example, allows us to resolve what we might call 'valuation problems'. Lewis offers an example of a valuation problem: suppose we have many merchants in a port town. Each wants to exchange his goods for some commodity he can in turn exchange for the goods and services he needs. It should not really matter to anyone what commodity they accept; only that that commodity will also be accepted by other merchants. No one wants to be stuck with a commodity they cannot use; this is a co-ordination problem (7). The problem is solved when we

introduce a coin of the realm. It is conventional that the coin of the realm will be accepted in exchange for goods and services, but the coin of the realm is only one proper co-ordination equilibrium amongst myriad possibilities.

This example is meant to lend itself to generalization: if the cases of signalling and money are not relevantly different from any other social institution (they aren't), then all status-functions are conventional solutions to co-ordination problems. But we might search for other arguments that don't rely on generalization. We might begin, as Lewis often does, by reciting a number of examples that reflect the diversity of the social world:

1. When I have passed the appropriate tests and paid the required fees, I receive my drivers' license. I am now authorized to drive.
2. When I have been arrested, charged, tried, and sentenced for a crime, I am held guilty of that crime. I am required to serve my sentence.
3. When I am caught parking inappropriately, I am given a parking ticket. I am required to pay the ticket.
4. When I run and win the most votes in a Republican primary race, I am named the Republican candidate. I am entitled to run for president.

As Searle aptly pointed out, even in the cases (1 and 3) where there is a physical representation of my social status (my drivers license, my parking ticket), it is still the agent that matters. Drivers' licenses and parking tickets have no real causal powers; they are merely representations, placeholders, 'status indicators' that indicate to other agents that I have had my conventional powers enhanced or reduced.

Furthermore, in each of these cases the agents' powers have been altered because she has met some conditions that we hold to entail those alterations. Winning a primary entitles you to run for president because we let people who've met that condition, and no one else, run for president. So, apparently, we construct social reality so that we can give order to the world we live in. People who meet certain conditions are given certain powers. But it is up to us which conditions will entail which powers, since it is within our power to change those entailment relations.

These entailments have not always existed. At some point, some person or group of people had to decide that meeting some condition would entitle them to some power. Perhaps there was a recurring problem of people exercising some power that seemed dangerous, such as driving when their limited skills made them a threat to themselves and others. Or perhaps there was a recurring situation of people meeting some condition that was problematic, such as leaving their vehicles in places that inconvenienced others. Populations experiencing such recurring problems might wish to find ways to ensure that such situations did not arise. But preventing such situations is clearly a co-ordination problem: we must have external, publicly accepted standards for what counts as qualified to drive, since if we were not to require such public acceptance then anything would count as qualified to drive. The problem would not be solved. The argument parallels Wittgenstein's claim that if there is no external standard for knowledge, there is no knowledge at all. You can't know you're qualified to drive unless you can compare your qualifications with some external standard. But developing external standards is a co-ordination problem. External standards are only valuable if everyone accepts the same standards for what counts as qualified. If

standards are not widely accepted, they are not standards: they are just some individual's beliefs about what counts as qualified, and she has no real justification for her claim that what she takes her qualifications to be really qualify her to drive. There are several possible sets of standards (as shown by the different standards in different countries), any of which would do. Our problem is solved when we manage to get everyone to agree on the same set of standards.

The same argument works for our other three cases. We need external standards for when a person is entitled to run for president, required to pay a parking ticket, required to go to jail, or is awarded any other conventional power (positive or negative). Setting external standards requires public agreement, and finding public agreement is a co-ordination problem. Once a particular agreement is in place, we have a convention since it is common knowledge that (1) almost everyone holds those standards; (2) almost everyone expects everyone else to hold those standards; (3) the psychological condition for being a co-ordination problem is usually met: at least when standards are being developed, people care more about finding agreement on some particular set of (one of the many minimally acceptable) standards, more than getting their first choice of sets; (4) almost everyone prefers to hold those standards, on the condition that almost everyone else holds those standards; (5) there is at least one other set of possible standards that is a proper co-ordination equilibrium that almost everyone would prefer to accept to, if almost everyone else were to accept those standards.

So we have the bare bones of a theory of the nature of social reality. We get social institutions when we need or want an external standard to regulate or organize some human behaviour. We need to be able to rule some behaviours out of order, such

as ungrammatical sentences, unlicensed drivers, and double parking, or else we won't be able to accomplish things that we want to accomplish (communicating, safe driving, free flow of traffic). So we might say that all of social reality is just the set of rules or norms that are conventionally arrived at as solutions to co-ordination problems and call some human behaviours illegitimate.

2.2.4. Another Argument, Suggestive of Things to Come

We can find another argument to this end if we accept one premise from Searle: that all pieces of social reality are status functions. First, I will show that all status functions are solutions to co-ordination problems; second I will show that those solutions are conventions. This shows that being a status function is sufficient for being a conventional solution to a co-ordination problem.

Searle defines 'status-function' as the conventional power afforded the wielder of some object as a product of our collective acceptance of the constitutive rule 'x (the object) counts as y'. At some point in the history of that constitutive rule, we must have been faced with a co-ordination problem: which object should count as y? We must necessarily have been faced with some choice of objects v, w, x, that we could potentially have counted as y. If there was never any such choice, then x simply is y and there is and never was any need for collective acceptance that x counts as y. By the same token, each of v, w, and x must have been proper co-ordination equilibria. Imagine if there was only one sort of thing, x, that we could use as an abstract medium of exchange. If this were the case, we have no need to collectively accept that x counts as money (see figure 2). Since x is the only thing that can be a medium of exchange, if we want to have a medium of exchange it has to be x. Furthermore, the suggestion that

there are other possible media of exchange that are not proper co-ordination equilibria is excluded ex hypothesi. If x is the only proper co-ordination equilibrium then again we have no need for collective acceptance in order to make x count as y. Each player has determinate reason to treat x as y even if the other players don't (because that's the best strategy). Remember from our definition of co-ordination problems, in games with only one proper co-ordination equilibrium, players need not consider what the other players will probably do. But when we are collectively deciding what should count as y, we must consider what the other players will probably count as y, or else our decision will not be collective.

Clearly I am appealing to some understanding of collective decision making that I have not argued for. But certainly any reasonable theory of collective activity must require that agents working collectively are in some sense working together. In problems with only one proper co-ordination equilibrium players need not work together. So, I conclude, the assignment of a status-function is a solution to a co-ordination problem.

Furthermore, the solution is necessarily a conventional one. The conditions for conventionality are clearly always met, just in virtue of the nature of the assignment of status-function. In order to generate a status-function in a population P, almost everyone in P must accept that x counts as y; almost everyone must expect everyone else to count x as y; the psychological condition for being a co-ordination problem is usually met: co-ordination is what really matters to most of the people (it matters more that we arrive at a proper co-ordination equilibrium than that an particular solution); almost everyone prefers to count x as y, on the condition that almost everyone else

counts x as y; and there is at least one other proper co-ordination equilibrium (w counts as y) that almost everyone would prefer to conform to, if almost everyone else were to accept that w counts as y. So, by Searle's definition, all status-functions are proper co-ordination equilibria and conventional solutions to co-ordination problems.

2.3.1. Conclusion

This shows that Searle and Lewis' theories of social reality are closer than they may have appeared on the first glance. In the next chapter I will compare their theories on several other dimensions, and argue that they differ on only one.

Chapter 3

3.1.1. Introduction

In this chapter I will compare and contrast the Searlian and the Lewisian accounts of social reality on several dimensions: the domain of the social world, the role of arbitrariness, the role of rules, the role of collective intentionality, the role of the Background and common knowledge, the role of language, and the role of negative reinforcement. I will argue that their views are nearly equivalent on nearly all of these dimensions.

3.1.2. The Role of Negative Reinforcement

The role of negative reinforcement is clear in Searle. We have to be prepared to censure those who break the rules of social reality, or at least feel that there is something wrong when someone breaks those rules. Furthermore, our learning of how to perpetuate the social world is not a matter of learning to follow rules, but rather a matter of learning to behave appropriately, and that learning would be much harder to come by if inappropriate behaviours were not negatively reinforced.

The Lewisian account agrees with Searle on this point, but we must be careful about how we state that agreement. Lewis has some interesting things to say about negative reinforcement in co-ordination problems. Consider this example: drivers face a co-ordination problem when they are deciding what side of the road to drive on. It doesn't particularly matter what side we drive on, as long as we're driving on the same side as everyone else. Suppose we arrive at the convention of driving on the right. If the highway patrol catches me driving on the left, I am punished. But, Lewis argues, these sorts of sanctions do not really affect our behaviour. If the sanctions coincide

with the co-ordination solution, then it is not the sanctions, but the fact that driving on the right is a co-ordination solution, that convinces us (for the most part, at least) to drive on the right. On the other hand, if the highway patrol were to sanction driving on the left, we would still prefer to drive on the right on the condition that everyone else does. Our preference for driving on the same side as everyone else outweighs any incentive the highway patrol could give us to drive on the opposite side. "The punishments are superfluous if they agree with our convention, are outweighed if they go against it, are not decisive either way, and hence do not make it any less conventional to drive on the right" (45).

But there is still room for negative reinforcement. In a situation with a conventional solution such as driving on the right, it is common knowledge that almost everyone expects everyone else to drive on the right. Furthermore, it is common knowledge that everyone around me has a reason to believe that I have a reason to believe that (1) conforming answers to my own preferences, (2) conforming answers to their preferences, and (3) they have reason to expect me to conform. So if I fail to conform, they may justifiably infer that I have acted contrary to everyone's interests as well as everyone's reasonable expectations. They will be surprised, perhaps upset, and will probably explain my actions discreditably, lower their opinions of me, and so on. They will rebuke and distrust me. "These are bad consequences, and my interest in avoiding them strengthens my conditional preference for conforming" (99-100).

So negative reinforcement makes a difference, but we don't need negative reinforcement in order to develop a social world. In all likelihood, we could get a social world off the ground without any sort of negative reinforcement, assuming that people

can adequately learn the appropriate co-ordination solutions without negative reinforcement. Indeed, perhaps the only real use either theory has for negative reinforcement would be in response to the question, "How is it that people in fact learn what behaviours are socially appropriate?"

3.1.3. The Role of Arbitrariness

Both theorists recognize the degree of arbitrariness in social reality. Searle recognizes that in many or even all instances of the constitutive rule 'x counts as y' there is more than one object that could conceivably be taken as the x-term. In the case of money, for example, he writes, "That objects can function as a medium of exchange is not a matter of convention but of rule. But *which objects* perform this function is a matter of convention" (CSR 49). There are some characteristics that we should want the x-term to meet: ease of transport and exchange, non-ease of counterfeitability, sufficient degree of durability, and so on. But no matter how long our list of characteristics gets, there will inevitably be several candidates that meet all of the conditions on the list. No list of x-term characteristics adds up to a y-term.

Lewis agrees with at least the latter part of the above quote from Searle. Since each solution to a co-ordination problem is only one of at least two possibilities (by nature of the definition of 'co-ordination problem'), there is necessarily at least one other possible conventional solution. Since co-ordinating matters to us more than arriving at any particular solution, which solution we actually arrive at somewhat arbitrary.

3.1.4. The Domain of the Social World

For Searle, the domain of the social world is all forms of power where sufficiently widespread belief that an object has a certain power is constitutive of the object having the power. The purpose of status-functions is to enhance or restrict the conventional powers of some agent. The Lewisian view agrees with Searle on this dimension. It claims that developing standards is a means of regulating or organizing agents' behaviours so that we can more easily reach our goals.

3.1.5. The Role of Language

For Searle, it is impossible for a non-linguistic community to generate a social world, since we can only call an x a y if we can call an x anything at all. But does it take a language in order to arrive at a conventional solution to a co-ordination problem? It would seem so, if common knowledge is to allow us to generate beliefs of arbitrarily high order about the agents we are interacting with. When we have common knowledge that x, I am entitled to infer that you believe that x, which entitles you to infer that I believe that you believe that x, which entitles me to infer that…and so on. It's hard to see how this complex pattern of inferences could occur in a species without linguistic representations.

But Searle's argument also seems to apply to Lewis' case. We suggested in the last chapter that Searle's claim that all social institutions are status functions is consistent with Lewis' theory, since all status-functions are conventional solutions to co-ordination problems. Since we have reason to think that the Lewisian account requires that social reality is impossible in non-linguistic communities, and there is a good reason to think that language is required not only for the generation of common

knowledge but also in order to treat some object (x) like it's something else (y), we can happily concede that the Lewisian approach requires that social reality can only exist amongst language speaking populations.

It might be objected here that the Lewisian approach can explain social reality without suggesting that members of any population treat some object (x) like it's something else (y). The Lewisian approach turns this object (paper) into that object (money) by making it common knowledge in a population P that everyone in P acts as though this paper were money (where such actions are a set of conventional solutions to a co-ordination problem). But as long as the members of P don't call that paper 'money,' they can have social reality without meeting Searle's linguistic requirement.

This objection is blocked by the common knowledge requirement. It is not even in principle possible to have a population with conventional solutions to co-ordination problems without language, since that population won't be able to have common knowledge. But this objection might be seen as demanding that if I want to see the Searlian and Lewisian positions as equivalent on this dimension, the Lewisian account should hold that social reality is impossible in a non-linguistic population for Searle's reasons. Searle's reason is essentially that you can't have constitutive rules without language *because the y-terms of constitutive rules cannot exist outside of a linguistic framework*, and without constitutive rules you can't have social reality.

In the next section I'll explain why Lewis might accept that you can't have social reality without constitutive rules. But I think I can also commit the Lewisian account to the italicized because-clause above. Most importantly, we couldn't say that a community that treated some stripe of paper like money without speaking a language

was actually using the paper as *money*. We might say that in this community the paper functioned *just like* money, or *as if* it were money, but we'd have no reason to think that it actually was money. If we properly hold that the members of this population do not understand (that is, it is not sufficient that the members are merely mute-they must have no linguistic comprehension) that their possession of the right stripe of paper counts as money, then we would be much better off explaining their exchanges of paper for goods by stipulating that they must find the paper intrinsically valuable. Saying that they use the paper as money implies that they understand all sorts of things that cannot be understood by non-linguistic beings.

I should be careful about one final point. It might be suggested that Lewis is offering his account of conventions in order to explain how people *get* language, so it should not assume that they already have language. Of course, I agree that for Lewis solving co-ordination problems does not presuppose a public language. I want to suggest that it does presuppose the capacity for language, such as mental representations in the absence of language.

3.1.6. The Role of Rules

For Searle social reality is a system of constitutive rules. Constitutive rules are rules that can bring social structures, such as basketball and socialist governments, into existence. Basketball exists precisely because we play basketball-that is, we accept and follow the constitutive rules for basketball. The same goes for all of social reality: money exists because we use money; governments exist because we elect them and treat them like governments. Searle holds that since producing social reality is a matter

of following rules, it is not a matter of upholding conventions[8]. He does not characterize the notion of 'convention' he has in mind, but it seems like Searle's usage is one of the many colloquial senses of the word.[9] Lewis is clear that his sense of 'convention' may not be coextensive with any colloquial sense (46).

As Searle looks dimly on the notion of a social world created by conventions, so Lewis looks rather dimly on the world of rules. He writes that "the class of so-called rules is a miscellany, with many debatable members. We might be tempted to try distinguishing several senses of the word "rule," hoping that one of them would agree with my definition of convention. I doubt that the project would succeed" (105). Nonetheless, we shall see that Searle's constitutive rules and Lewis' conventions, though hardly coextensive, share important properties that suggest a resolution Searle's concerns.

The Lewisian account of social reality says that social reality is a set of conventional solutions to co-ordination problems. To find out whether there is a serious difference between Searle and Lewis on this point, we might ask whether the Lewisian claim might be recast in Searle's language of constitutive rules. Let's tell the Lewisian story from the beginning, attempting to substitute the appropriate terms.

People at times face recurring problems that they wish would stop arising. Examples of such situations are accidents caused by unskilled drivers, obstruction of traffic due to unrestricted parking, or uncertainty about what sort of currency to accept in exchange for goods. We can prevent these sorts of problems by developing external,

---

[8] As I quoted him earlier, "That objects can function as a medium of exchange is not a matter of convention but of rule. But *which objects* perform this function is a matter of convention" (CSR 49).
[9] For some comments on the tangled colloquial senses of 'convention,' see Millikan's "Language Conventions Made Simple" note 2.

objective (i.e., collectively accepted) standards that regulate the behaviour of agents. Once we all agree on some characterization of the standards, we follow those standards. We require that anyone who wants to drive legitimately pass a certain battery of tests. Passing the test counts as getting a license, and having a license counts as being entitled to drive. The institution of licensing exists because we play by its constitutive rules.

Now we might ask, does anything said here contradict the account of social reality offered in the last chapter? More pointedly, does the appeal to constitutive rules preclude the possibility that we are following conventions in Lewis' sense? It does not. Lewis even gives us an example of a game called Jotto, which is any set of behaviours that conforms to these (constitutive) rules:

1. There are two players.
2. Each player chooses a five-letter word, which is kept secret.
3. Players take turns guessing five-letter words. After each guess, the non-guessing player tells the guesser how many of the letters in the word guessed were in the secret word.
4. When one player guesses the other's secret word, the game is won.

These rules are conventions according to Lewis' definition (104). When we are choosing the rules of a game, we do not care (within a certain range) which set of rules we use; only that all players play by the same set of rules. So, describing a pattern of behaviours as following some constitutive rules does not entail that the same a pattern of behaviours cannot be described as following some conventions. Indeed, Lewis takes constitutive rules to be a proper subset of conventions. There may be other conventions of the game that cannot be said to be part of the official rules of Jotto, though they may

count as informal or unwritten rules. How is it to be decided which player goes first? Which types of words are admissible? These sorts of rules tend to vary across groups of Jotto players, while if some member of rules 1-4 above were to vary we could not say that those players were playing Jotto. Lewis captures the difference by saying that the variable, informal sorts of rules are not fully rules but merely conventions.

It's important to note that there's no reduction from Searle's notion of constitutive rules to Lewis' talk of conventions. The rules of Jotto are all conventions, and playing by those conventions counts as playing Jotto. So the 'counts as' locution is necessarily part of the Lewisian theory as well: without it the Lewisian cannot take the step from 'this convention is a solution to a co-ordination problem' to 'this is a game.'

Still, Searle claims that if we are following constitutive rules, we are not following conventions; Lewis claims that following constitutive rules is just one way of following conventions. It is tempting to say that the difference is a terminological one, especially since Lewis' definition of 'convention' is rather technical and may (as Lewis admits) diverge from ordinary usage. Also, since Searle does not characterize his notion of convention, he is likely using it in one of its ordinary senses.

3.1.7. The Role of the Background/Common Knowledge

The Background is a set of bits of know-how that is developed through a non-rule based process of trial and error[10]. Our ability to generate and interact with the social world is produced by these background abilities. Remember our characterization in chapter one of the three requirements for the Background to cause the preservation of

[10] Of course, the Background is a part of our neurological wiring, and our neurological wiring is in part a product of evolution and in part a product of our learning as individuals. Following Dretske, we might say that there are two processes of trial and error at play in the development of the Background: one phylogenetic (evolutionary trial and error) and one ontogenetic (a product of individual learning).

a social institution. First, we must born into a social world where the social institutions in question are in some sense useful. Second, because social institutions are both useful and ubiquitous, we learn rather unwittingly and from an early age how to use them. The learning is unwitting because it does not take place in a principled manner; rather, we acquire a bike-riding style (Background) know-how. Third, because of that know-how, caused in us by rules in a neither intentional nor physical way, we continue to navigate the world in a way that keeps the system going. We keep following the constitutive rules for that social institution, which is equivalent to keeping the social institution alive for another generation.

Compare this with the requirements for the generation of conventional solutions to co-ordination problems. Remember that it must be common knowledge that the five conditions for being a convention are met, and that we have common knowledge that x in a population P if and only if there is a state of affairs A such that:

A. Everyone in P has a reason to think that A is true;

B. A indicates to every individual in P that all members of P have a reason to think that A is true;

C. A indicates to every individual in P that x.

State of affairs A is then called the basis for common knowledge. But we might ask whether it is plausible to relate the basis for common knowledge A to our Background abilities to produce behaviour that is consistent with our social institutions.

Remember that the Background can be described as a set of rules that influence our behaviour, though through neither intentional nor physical causal pathways. We do not explicitly know these rules; they are rather like the rules for riding a bike. We

cannot say what they are, but we know very well how to follow them. So we can say that we know the rules, but it is know-how, not know-that. But suppose someone were to accurately codify the rules for riding a bike, or the rules we follow when we apply social institutions. Then presumably, everyone party to a social world, if presented with such a codification:

A. would have a reason to believe that the codifications were accurate (they should coincide with everyone's experience);

B. would have a reason to think that anyone able to enjoy the social institution would have a reason to think that the codifications were accurate;

C. would be informed that the rules as codified are in fact the rules that allow each individual to ride a bike or apply the social institutions.

Before going further, I'd like to distinguish two questions about the social world. There are questions about how we initially create social institutions, and then there are questions about how those institutions, once created, continue to exist. Lewis' common knowledge is a part of his apparatus in answering the former question. Searle uses the Background to address the latter issue. So, though we have seen an interesting parallel between the Background and common knowledge, there is an important difference. Common knowledge is an intentional phenomenon, and plays a role in an intentional process (arriving at conventional solutions to co-ordination problems), while the Background is by definintion pre- or nonintentional states that, amongst other things, allows for the nonintentional perpetuation of the social world. The two are occupants of different functional roles in different sorts of theories. But each theory needs something to fill each functional role in order to be a complete theory of the

nature of the social world: Searle needs to tell us how we manage to arrive at the right collective intentions when we are sculpting the social world, and the Lewisian needs to tell us how we perpetuate the social world unintentionally.

It is not part of Lewis' project to say how the social world is perpetuated. He tells us how language as a conventional phenomenon can get off the ground without presupposing language. So, giving an account of how the social world is perpetuated requires going beyond Lewis' work. But the thesis of the Background is not excluded as a possibility by Lewis' work in Convention, and the parallel drawn above between the Background and common knowledge suggests that there is some common spirit between the two notions. The Background offers a complimentary means of completing the Lewisian project.

Searle's theory is similarly complemented by Lewis' notion of common knowledge. When we are inventing a new game, or forming a new club, or constructing any new social institution, we have to get the right group of people to accept the right constitutive rules. The players have to accept the right rules of the game, the members have to agree on a constitution for the club, and so on. As we shall see in the next several sections, members of the collective have to individually intend to collectively construct their game, club, or whatever by means of individually accepting the rules of the game, club, or whatever. But in aiming to accept 'the rules of the game' as the rules of a game that is not yet invented, each member must be striving to accept the same rules as each of her counterparts-otherwise we will end up with several games, or no game, but not one game.

So it looks as though Searlean social constructors, in their initial creations of social reality, will have Lewis-style common knowledge. If we are to construct a game with G rules R1, R2, and R3 then it must be common knowledge that everyone who wants to play G should accept R1, R2, and R3 as the rules of G, since there is a state of affairs A (that if everyone who wants to play G doesn't accept R1, R2, and R3 as the rules of G, they won't be able to play G) such that:

A. Everyone has a reason to think that A is true;

B. A indicates to everyone that everyone has a reason to think that A is true;

C. A indicates to everyone that everyone who wants to play G should accept R1, R2, and R3 as the rules of G.

So, intentional Searlean social constructors have Lewis-style common knowledge.

3.1.8. The Role of Collective Intentionality

In Searle's theory collective intentions are required to generate social reality. According to him it is our collective acceptance or belief that someone or something has some conventional powers that *constitutes* or is *equivalent to* that thing or person having those powers. Take away the collective intentions, and the social institution disappears. Something similar can be said for the Lewisian account. In any case of the generation of social reality in a Lewisian world, players want to arrive at a co-ordinated solution more than they want any particular solution. They all want everyone to be aiming at the same co-ordination equilibrium. Take away the desire to aim at the same equilibrium, or even just take away the fact of aiming at the same co-ordination equilibrium, and we will no longer have a conventional solution to our co-ordination

problem. Without a conventional solution, we will not be generating social reality on the Lewisian account.

Lewis never talks about collective intentionality. But I am suggesting a parallel here between the role played by collective intentionality in Searle's theory and the role played by aiming at the same co-ordination equilibrium in Lewis'. The two notions are different, especially given Searle's anti-reductivist views about collective intentionality. But since they play the same role in two theories that are beginning to look more similar than they initially appeared, they may give us a criterion for choice between the two theories. So, in the next section of this chapter, I will examine carefully Searle's notion of collective intentionality. Then I will say clearly what it is that plays the same functional role in the Lewisian theory. Ultimately we will compare the two entities in choosing between the two theories.

3.2.1. Searle's Notion of Collective Intentions

Developing an adequate understanding of Searle's notion of collective intentions requires first that we develop some of his apparatus. In particular, we need to know something about his general account of intentionality.

3.2.2. Searle on Intentions

Consider three cases of human motion:

1. Your hand rises because of stimulation of electrodes implanted in your brain;
2. Your hand rises because you've decided that now is the appropriate time to ask the question you've been waiting to ask;
3. Your hand rises as a reaction to an obviously false and insulting point.

In the first case, there is nothing intentional going on; the person whose hand moves cannot be held responsible for the movement. The second case is a full-blown intentional action. You have an interesting question, and decide that when the appropriate time comes, you will raise your hand. This is what Searle calls a 'prior intention'. Every prior intention has conditions of satisfaction that include a particular motion and a particular cause of that motion. If I intend to raise my hand when the time is right, and when the time is right my hand goes up because of stimulation of electrodes implanted in my brain, then my prior intention is not satisfied. In order to be satisfied, my prior intention must play some causal role in producing the motion: namely, that I produce the motion in order to satisfy that intention. Mechanically speaking, the prior intention must cause an intention-in-action which causes the motion. Prior intentions extend through time: they are intentions we have now to do something later. Intentions-in-action, on the other hand, are not temporally extended from a present intention to a future act. They are a constitutive element of any action. An intention-in-action is an intentional state with an appropriately caused *motion* (ie, a motion caused by that intention-in-action), rather than action, as its conditions of satisfaction. Only satisfied intentions-in-action with the appropriate causes can satisfy a prior intention. A satisfied intention-in-action is by definition an intentional action, and since prior intentions can only be satisfied by satisfied intentions-in-action (i.e., by intentional actions), satisfying prior intentions requires intentional action. Intentions-in-action are the intentional states we need to have in order to move right now, rather than to plan to move later.

The third case above is an example of an intentional action without prior intention. We still have an intentional action, but since there is no premeditation there is no prior intention. There is merely intention-in-action, and that intention-in-action is satisfied because the hand goes up because of that intention-in-action. That means that the intention-in-action is satisfied (we have a motion with an appropriate cause), which means that there is an intentional action.

In short, Searle's account of intentionality says we have a premeditated intentional action when we have a satisfied prior intention. But prior intentions get satisfied by causing intentions-in-action which are in turn satisfied. Intentions-in-action get satisfied by causing bodily motions. Satisfied intentions-in-action simply are intentional actions. If any particular intentional action satisfies some prior intention, then the action was premeditated (Intentionality, chapter 3).

3.2.3. Searle on Collective Intentions

What, then, is a collective intention? Searle answers this question by analogy with a special sort of individual intention-in-action. Sometimes our intentions-in-action take the form "achieve-B-by-means-of-A" (Collective Intentions and Actions, 412). So, for example, I might have an intention-in-action to fire this gun by means of pulling its trigger. This intention, says Searle, is satisfied if and only if the intention causes it to be the case that the trigger pulls, which causes it to be the case that the gun fires.

Collective intentions are to be understood as one of these sorts of complex intentions-in-action. So, to use an example of Searle's, if we are collectively making a hollandaise sauce and my role in that collective activity is to slowly pour in the ingredients while you stir constantly, the content of my intention-in-action is,

'collectively make the sauce by means of individually pouring in the ingredients'. The intention-in-action is satisfied if and only if that particular intention causes the ingredients to get poured in, which in turn causes the sauce to get (collectively) mixed.

This structure of collective intentions is meant to show that collective intentions cannot be reduced to what individuals can intend to do themselves. There must be some reference in the structure of the intention to what the collective intends to do as well. If we try to reduce collective intentions to individual intentions, we end up with only individual acts and no sense of the part-whole relationship of each individual act. To illustrate this, Searle offers the example of several park goers suddenly caught in a rainstorm, each individually rushing towards a centrally located shelter. The non-reducibility of collective intentionality becomes clear when we contrast this case with a troupe performing a modern dance that replicates the events of the previous case. If collective intentionality were reducible to individual intentions, even as supplemented by beliefs about what others are doing, we would not be able to distinguish between these cases. But since there is a clear distinction-there is a collective action in the second case, but not in the first-collective intentionality cannot be reduced to individual intentionality.

Searle says that there are two sources of the difference in these two cases. First, in the former case, the intentions of the individuals could be expressed without reference to what others are doing or intending to do, which is not true in the latter case. Second, and we will return to this point shortly, each individual's intentions to do *A* in the second case are in a certain sense *derived* from the collective intention to do *A*.

3.2.4. Searle Against Reductive Collective Intentionality

Searle argues briefly (and, admittedly, inconclusively) that reductive accounts of collective intentionality always admit of counterexamples. He offers this argument: the best reductive account of collective intentionality admits of a counterexample, so it is plausible to conclude that they all will. The account he considers is offered by Raimo Toumela and Kaarlo Miller. They argue that Pat, as a member of a group with a collective intention, collectively intends to do *X* if:

1. Pat intends to play his assigned role in doing *X*.

2. Pat believes that success is possible: in particular, he believes that the other members of the group will (probably) fulfill their assigned roles.

3. Pat believes that everyone else in the group has the belief required by condition 2.

Suppose, then, that Pat is one of a group of potential businesspersons entering a business school. They learn about Adam Smith's theory of the invisible hand: that the greater good is best advanced by pursuit of selfish ends. They are all convinced by Smith's arguments, and furthermore know that all are convinced, and each individually resolves to advance humanity by that selfish pursuit. We might then say that:

1. Pat intends to play his assigned role in promoting the greater good (ie, he intends to pursue his selfish ends).

2. Pat believes that success is possible: in particular, he believes that the other members of the group will (probably) fulfill their assigned roles (ie, they will pursue their selfish ends).

3. Pat believes that everyone else in the group has the belief required by condition 2.

So, Pat satisfies the conditions for having collective intentionality, but in a case where there clearly is no collective intentionality. The whole idea behind the invisible hand is that it works best when there is no collective intentionality. Of course, if the students at some point had gotten together and decided to each go about pursuing their own ends to the betterment of the community, then we would have a case of collective intentionality. But that is not the case in this example.

Searle then examines the concern that we might adjust our account of 'play an assigned role' so as to preclude such counterexamples. If we adjust the notion of playing an assigned role so that it means doing something towards a collective end, then we build the notion of collective intentions into our explanation of collective intentions. Any such explanation of collective intentionality would fail as a reduction because it fails as an explanation at all: it is viciously circular. On the other hand, if we do not build a notion of collective intentionality into our (reductive) account of collective intentionality, then it will admit of the sort of counterexample outlined above. This is the reductivist's dilemma.

The reason why collective intentions are not some summation of individual intentions is because no such summation can allow any real understanding of cooperation. These summations, even when supplemented by beliefs and mutual beliefs of the agents concerned, cannot entail the presence of an intention to cooperate. Because of this failure, there will always be cases where the intentions and beliefs are

present but the collective intention is not. In this thesis I am introducing the Lewisian notion of co-ordination in order to overcome this argument.

Let's summarize Searle's non-reductive account of collective intentionality. A collective intention is an intention inside the head of an individual to achieve a collective end by means of an individual contribution. Presumably, every individual in the collective has an intention to achieve the same collective end by means of various individual contributions. So, Searle shows a flaw in the reasoning of many theorists who want a reductive account of collective intentionality. They think, says Searle, that since an individual's intentions can only cause that person's actions, those intentions must refer only to the actions of that individual. Indeed it is the case that an individual's intentions (call her A) cannot refer to the actions of another individual (call him B), since if B's actions were determined by A's intentions it would not be clear in what sense they were still B's actions. However, the set of actions are not exhausted by the actions of individual agents; there are also actions of collectives. And the intentions of individuals can, and indeed must (if we are to explain collective action without appealing to a Hegelian 'world spirit' or unworkable reductions) make reference to the actions of collectives (CSR 25-26). Thus we have the 'by means of' structure of collective intentionality. I intend to participate in collective act *A* and can best contribute to the success of *A* by performing individual act *B*. So the content of my intention is 'to perform collective act *A* by means of individual act *B*.'

3.2.5. A Helpful Notation

When someone asks two cooks preparing a meal together, "What are you guys doing?" there are two sorts of answers, each of which satisfies one of two important

intuitions we all share about collective intentionality. They might say, "We are preparing the sauce." Or they might say, "I am stirring" and "I am pouring." The two intuitions are these: in the first case, collective action is a real phenomenon. It really is the case that we do things together. The second intuition is that all actions are carried out by individuals. The question at hand is whether the intentions that lead to collective actions make necessary reference to the (real) collective entity, or whether the content of those intentions can be expressed in fully individualistic vocabulary.

Searle offers a notation that makes this question much easier to address. The relevant features of any intention can be captured in this structure:

x (x causes: y) CAUSES: Y

The x is to be replaced with a type of intentional state, such as prior intention or intention-in-action. The lower case letters inside the brackets represent the mental component of the intentional state and specify the conditions of satisfaction for that intentional state: that that intentional state causes some state of affairs propositionally described. The upper case letters outside the brackets represent physical states of affairs in the world: the intentional state causes the pattern of changes in the world specified by the propositional content of the mental component.

So, consider the third example of human motion offered at the beginning of this section: your hand rises as a reaction to an obviously false and insulting point. I said earlier that this is a case where we have an intention-in-action (i.a.) with no prior intention. So the relevant features of this intention can be captured thus:

i.a. (this i.a. causes: my arm goes up) CAUSES: MY ARM GOES UP

So the type of intention is an intention-in-action, the mental component is that that intention-in-action causes the arm to go up, and the intention is satisfied just in case the intention-in-action in fact causes your arm to go up.

Since Searle takes himself to have established that talk collective intentions cannot be eliminated in favour of talk in individualistic idiom, he wants to make sure there is an adequate way to represent collective intentions in this notation. So, in the case of collective sauce making, where we have a collective intention in action (c.i.a.), what's going on inside the head might look something like this:

c.i.a. (this c.i.a. causes: sauce gets made) CAUSES: SAUCE GETS MADE

But this formulation neglects the by-means-of structure of collective intentions. We know that collective intentions have the by-means-of structure because our common understanding of the role of individuals in group actions is reflected when we say, "we are making the sauce by means of me stirring," given that that is equivalent to "stirring is my part of our making the sauce."

Now, as in the case of firing the gun by means of pulling the trigger, there are not two actions, one collective (making the sauce) and one individual (stirring the ingredients). Rather, stirring the ingredients is in part constitutive of making the sauce. So we can more adequately capture what's going on with this form:

i.a. collective B by means of individual A (this i.a. causes: A-stirring causes: B-sauce is made) CAUSES: STIRRING CAUSES: SAUCE IS MADE.

This representation requires that there is only one intention, and hence only one action that is in part constitutive of the collective action. The individual action is derived from

the collective action because the collective end (getting the sauce made) is what matters, and the individual action is a means to that collective end.

3.3.1. Intentional Contents in Lewisian Social Constructors

In order to compare Lewis with Searle on this front, I'll ask what has to be going on inside the heads of people who are constructing social reality by finding conventional solutions to co-ordination problems.

Lewis' definition of 'convention' requires that everyone participating in the convention (1) knows that everyone usually does their part in the co-ordination solution; (2) expects everyone else to do their part; (3) is aiming primarily at co-ordination, and only secondarily at their preferred co-ordination solution; (4) prefers to do her part on the condition that everyone else does; (5) has at least one other option that she would prefer to do on the condition that everyone else was going to do that action. This implies several things about the agents' intentional states, but most importantly she intends to play her part in whatever co-ordination solution is chosen by the others she is co-ordinating with. What, then, is the content of this intentional state? More importantly, does that content necessarily refer to the act of a collective, or can it be stated in purely individualistic idiom?

Let's take a few examples of what that content might be. For brevity, I'll leave out the upper-case clauses that represent the satisfaction of the intentions in action. Consider:

1. i.a. (this i.a. causes: I do x on the condition that x is my part of the co-ordination solution everyone else involved is aiming at).

2. i.a. (this i.a. causes: I do x) (suppose that x is my part in reaching the co-ordination equilibrium, but I don't know that).
3. i.a. collective B my means of individual A (this i.a. causes: A-stirring causes: B-sauce is made)

Description number 3 is Searle's non-reductive view. My intention makes an ineliminable reference to the collective action to which my individual action contributes. Trying to explain away the reference to the collective action by appealing to my beliefs about what others are doing will inevitably leave something out, and thus not succeed as an explanation.

Description number 2 is an unsatisfactory description of the content of my intention. If we take this as my intentional content we lose any sense of the role of co-ordination. Remember that my intention should be conditional: in a conventional solution to a co-ordination problem, I only intend to do x on the condition that everyone else plays their role in reaching that same co-ordination equilibrium. Since this description does not capture this conditionality, it is inadequate.

Description number 1 does capture that conditionality. The more interesting question is whether it makes ineliminable reference to the actions of the collective. The answer is no. Searle's arguments about what is left out when we try to augment our individual intentions with beliefs about the intentions of others do not leave anything out in this case. We still grasp conditionality, and we still have the sense of trying to work together or co-operate with other members of a group. But our intentions need not make any reference at all to the intentions of the group; we can talk solely about the intentions of each individual and not lose any sense of co-operation between the

individuals. They are co-operating because each individually wants primarily to co-ordinate, and each individual is prepared to do their part to reach a co-ordination solution. Their action can be fully described on two levels: we can say that they are collectively doing y ('we are making a sauce') or individually doing u, v, and w, their respective parts in reaching the co-ordination solution everyone is aiming at.

So, it looks as though the Lewisian account may be able to get by without non-reductive collective intentionality. Given the similarities we found between this theory and Searle's at the beginning of this chapter, it looks like that may be the only significant difference between the two. Remove the non-reductive collective intentionality from Searle's account, and you're left with the Lewisian view. This means that we can take the issue of whether there is non-reductive collective intentionality to be decisive between the two. We can treat it as a criterion for choice.

Chapter 4

4.1.1. Introduction

We need now to examine the nature of collective intentionality and the role it plays in the theories at hand, so that we can choose between them. First I'll consider some arguments a Lewisian might make contra Searle; then I'll look at some objections an anti-reductivist might make to the Lewisian.

4.1.2. A Lewisian Response to Searle: The Burden of Proof

In the last chapter we saw that Searle's main argument is that the best reductive account of collective intentionality fails, so we can plausibly conclude that all such attempts will fail. This is essentially a burden of proof argument. It challenges us to offer a reductive account of collective intentionality that neither builds in a notion of collectivity nor neglects the role of co-operation. I find it strange that Searle examines the Tuomela-Miller view, but does not consider Lewis' view or any view at all based on co-ordination. This is especially odd when we realize that the beliefs and mutual beliefs ascribed to members of collectives in the Tuomela-Miller account are also ascribed on the Lewisian account be the appeal to common knowledge. The Lewisian account simply adds the claim that players must be trying to co-ordinate, which adds the element of co-operation that Searle thinks is missing from reductive accounts. The Lewisian account thus makes a stronger claim than the Toumela-Miller account. Since Searle wants to examine the best reductive account of collective intentionality, and it seems like co-ordination is the best candidate for a non-collectivist account of co-operation. So, we might rather rest the burden of proof in Searle's shoulders by suggesting that he has not considered the best candidate for a reductive account.

4.1.3. A Lewisian Response to Searle's Dilemma

What might Searle say if we were to offer him this burden? Most likely he would suggest that it was no burden at all, and argue that the Lewisian account of collective intentionality falls on one of the horns of his dilemma. Let's consider each horn in turn.

I think the easier of the two horns to defend is the suggestion that the Lewisian account fails to give us a real sense of co-operation. Lewis formulates his position carefully so as to ensure that he does give us a real sense of co-operation. When we are faced with co-ordination problems, by definition the psychological condition is met: our priority is to find a co-ordinated action, not to find any particular co-ordination (or otherwise) solution. What we are trying to do is work with the other players (as individuals, not necessarily as a group) in order to arrive at a co-ordination solution. We are trying to find a way to co-operate.

Does the Lewisian account incorporate a notion of collectivity? In the last chapter I gave some reasons to think it doesn't. The content of my intentions when I am faced with a co-ordination problem is, "i.a. (this i.a. causes: I do x on the condition that x is my part of the co-ordination solution everyone else involved is aiming at). There is no need to refer to the intentions of the group; we need capture only the intentions of the other players of the game. Furthermore, nowhere in the presentation of the Lewisian account was there any need to introduce groups as entities; the entire position was carefully formulated to avoid necessary reference to groups as entities. All of the theoretical terms are or can be defined in terms that need not refer to groups as entities. Thus, for example, the use of 'population' when we defined 'common knowledge' and

'convention' could be substituted with 'aggregate of individuals' without loss of meaning[11].

These considerations suggest that the Lewisian account may be protected against against the kind of counter-examples Searle launches against the Tuomela-Miller account. His counter-examples turn on the insufficiency of reductive characterizations. We might try to conceive such a counter-example in Searle's style to the Lewisian approach. Searle's well-intentioned businesspersons will not do the trick, since they are not faced with a co-ordination problem. Each does not prefer to conform on the condition that the others do, since each prefers to conform no matter what; each has a decisive reason to go about pursuing her own selfish ends no matter what the others do. Furthermore, it seems likely that if we alter the counter-example to make it into a co-ordination problem it will no longer be a counter-example. If we manage to give each businessperson the right sort of conditional preferences then we'll have given them collective intentionality. Searle hasn't given us a counter-example, or an obvious way to construct one. The Lewisian account appears sufficient for collective intentionality. So not only is the burden of proof with Searle, but it is prima facie a real burden.

4.1.4. Why Would Anyone Want a Reductive Account of Collective Intentionality?

If I have been at all persuasive so far, then we are faced with two theories that are more or less identical in all respects, with the exception of their accounts of the nature of collective intentionality. One (the Lewisian) requires that there are intentions that belong to individuals that refer to the actions of individuals. The other requires

[11] Of course, if that aggregate does not form a social group we might doubt that the conditions for convention or common knowledge will ever be met, but that is only because social groups are the subset of the class of aggregates of individuals wherein those conditions are most easily met. This point will come to the fore in my discussion of Gilbert, below.

both such individual intentions (for when agents act alone) and a further sort of non-reductive collective intention. Occam's razor suggests that we should take the former, as does the burden of proof. Collective intentions do not fit snugly into received views in theory of action and rationality; the Lewisian account does, which affords it greater inter-theoretical integration. The Lewisian account offers a reduction of collective intentionality, and when such reductions are forthcoming they increase our explanatory power. In the remainder of this chapter I'll review some reasons that people have held that reduction in this case will not be forthcoming. We'll find that there are real obstacles to such reductions, and serious reasons to think that Lewis' project at least cannot succeed. I'll start by reviewing some objections that I think Lewis can satisfactorily answer, but the responses will grow less and less likely until we arrive at problems that are genuinely insurmountable.

4.2.1. Methodological Individualism?

In this section I will consider whether Lewis is committed to an untenable methodological individualism. Since his account of convention is stated in individualist idiom, he is at least open to the suggestion that all social groups are constituted by the acts of singular agents[12]. In this Harold Kincaid, following Hellman and Thompson, distinguishes two claims. First, social groups are exhausted by individuals because once you've accounted for all of the individuals and their interrelations there's nothing left over. Second, social groups are determined by individuals, since once all the facts about individuals and their interrelations are known, there is nothing left to know about

[12] A singular agent is one whose acts are determined by their personal preference schema, i.e., their own desires broadly construed, rather than the desires of any group of which they are a part. See Gilbert, and below. The issue is important since if there are acts determined by group preference schemas rather than individual preference schemas, then we have reason to think complete reduction from the social to the individual is implausible.

social groups. So, any two social domains with the same individuals in the same individual relations will have all of the same properties. The social world supervenes on the individual world.

MacDonald and Pettit argue that these commitments bind their bearers to saying that theories that refer only to individuals (to the exclusion of social entities) are capable of explaining all social phenomena. Their argument is such: first, the types of events that social theoretic explanations target are also targeted by the 'orthodox conception of agents'. In other words, folk psychology (the orthodox conception of agents) claims to have the conceptual resources to explain everything in the domain of social theories, since it explains all intentional acts. Second, explanatory collectivism (the view that social theories can explain things that individualist theories cannot) must deny the orthodox conception of agents in holding that that conception cannot explain all it aims to explain. From their third premise, that the orthodox conception of the agent is undeniable, it follows immediately that explanatory collectivism is false.

I'll quickly review their arguments for the undeniability of the orthodox conception. First, the orthodox conception is both irreplaceable and unrevisable. It cannot be replaced by any other explanatory theory such as an advanced version of our present neurology because without ascribing beliefs and other propositional attitudes to others we cannot interpret the semantic content of their utterances[13]. It is unrevisable because of the nature of intentional explanation: such explanation is non-nomothetic. It does not involve trying to find the appropriate explanatory principles that accord with the evidence. We have already determined the principles. They are held fast, and when they are challenged we find interpretations of the facts that eliminate the challenge

[13] The argument is essentially Davidson's. See "The Method of Truth in Metaphysics."

(Semantics and Social Science 95). Undeniability and unrevisability make the orthodox conception impregnable: its undeniability ensures that there is no better explanatory theory, and unrevisability ensures that once it's put in place as an explanatory hypothesis it attains an unquestionable status. So, not only do we need the orthodox conception-once we have it we can't get rid of it. So, any position (such as eliminative materialism or explanatory collectivism) that requires we deny the orthodox conception must be rejected.

Kincaid argues that this consequence of individualism is unbecoming. If individualist theories are capable of explaining all social phenomena (i.e., all things explained by social theory), then individualist theory should be able to reduce social theory. This is an implausible claim, since there are really only two ways for such a reduction to occur, neither of which appear very promising.

The first unpromising route is old-fashioned reductionism. Someone who holds that theories that refer only to individuals are capable of explaining all social phenomena must hold that individualist theory can do all the work done by our social scientific theories. For each event or entity in a complete social theory, there must be an expression in individualist terms. We need to find individualist equivalents to social-theoretical types. This is traditional 'bridge-law' reductionism, and Kincaid gives us three reasons to think that such a reduction is implausible in this domain. First, social phenomena are multiply realizable: different bureaucracies may supervene on many different arrangements of people, and even a single bureaucracy may supervene on several different arrangements of individuals. So one direction of the biconditional in the bridge-laws relating social and individualist phenomena seems to fail. The other

direction also fails, since a single event individualistically described might have different social significances depending on how we fill out various social details. Thus, two men shaking hands might be meeting one another, agreeing to a business deal, or marrying their children, depending on the social context. These two states of affairs cast doubt on the possibility of establishing bridge principles. A third criticism of the reductivist program aims at the likelihood of establishing a real reduction even if we have bridge principles. Individual social theory types such as 'teacher' or 'inmate' presuppose a lot of institutional context that refers to social entities for which there appears little possibility of expressing in individualist theory: you can't have a teacher without schools, and you can't have inmates without prisons, judges, lawyers, courts, and so on. Since any particular 'reduction' introduces a host of new entities to reduce, writes Kincaid, the prospects for reduction are unfavourable.

But there is a second possible sense of the claim that individualist theory suffices to explain all social phenomena. Suppose that the individualist means only that all social event or entity tokens can be fully explained in individualist terms. Kincaid responds that even if every actual social token can be reduced to some combination of individuals, there will still be work that can only be done by social theory. First of all, such token-token reductions may not support the same counterfactuals as social theory. If such reductions take the form of disjunctive definitions (the Canadian federal bureaucracy is combination A at time t1 or combination B at time t2…) then the term 'bureaucracy' in the counterfactual, 'if John A. MacDonald had lost the election, Canada would have developed a very different bureaucracy' has no reference.

But there is a more interesting problem with this approach to reduction. It turns on Kincaid's rather reasonable criteria for 'sufficient explanation' or, in his terminology, 'full explanation'. An explanation is always in part an answer to a 'why' question. Every such 'why' question determines both a contrast class (Why does the Earth go around the Sun - rather than staying put, rather than the other way around, rather than not existing at all?) and a relevant answer-type (Why did John go to the fridge? Because he wanted a glass of water, because of the angles of force applied by his muscles and the friction between his feet and the floor, and so on). A purported answer to a 'why' question is a full explanation if and only if it satisfies the contrast class and answer type determined by the question. But individualist theory as a token-by-token description of actual social phenomena cannot fully explain social phenomena in this sense. The relevant answer-type for the question, "why is this bureaucracy so inefficient?" might need to refer to laws that entail that all bureaucracies with certain characteristics are inefficient: this is a type-A bureaucracy, and type-A bureaucracies are always inefficient. Such laws cannot be expressed in the sort of individualist theory at hand, since laws cannot be expressed in such a limited vocabulary. Laws relate types of things, and the individualist theory at hand can only talk about tokens. Therefore, individualist theory as a token-by-token description of actual social phenomena cannot fully explain social phenomena.

So, it appears, reductive views would commit Lewis to a conditional with a true antecedent (the social supervenes on the individual) and a false consequent (individualist theory sufficiently explains social phenomena). To combat this charge I will argue that Lewis is not committed to the conception of sufficient explanation that

Kincaid has in mind. Indeed, I think that this objection tries to commit reductionists to much more than they in fact need to be committed to. Reductionists do not need to claim that there is no work to be done by social theories. Sociology, anthropology, economics, and so on make valuable contributions to our understanding of the way the world works, and it is truly not the case that the work of those theories can be performed in a lower-level individualist theory. But all that exists, all of the values of our variables, are individuals.

This evasion of Kincaid's objection amounts to saying that Lewis might only affirm the antecedent of the problematic conditional (if the social world supervenes on the individual world, then individualist theory is sufficient to explain all social phenomena) in question and deny its consequent. If this is the case, it may appear that I should be able to show where MacDonald and Pettit's argument for the entailment of the consequent by the antecedent goes wrong. But though I do have some reservations about their arguments, I think more relevant problem is that Kincaid neglects one aspect of MacDonald and Pettit's apparatus.

They hold that social theory is not explanatorily autonomous of individual theory since social entities do not enter into causal interactions in ways not captured by individualist theories. However, they also hold that social theory is expressively autonomous of individual theory, since we may be able to formulate truths in social theory that cannot be formulated in individual theory due to the greater conceptual resources of the former. I suggest that Kincaid's example turns on an instantiation of expressive autonomy. Social laws are not causally efficacious. The sorts of behaviours governed by social laws are *described* as law-like, but the laws themselves have no

causal role in generating the behaviour in question. There is no entity that is trying to fulfill social laws. Just because individualist theory cannot capture the law-like relations between tokens of social events does not mean that individualist theory cannot explain all of the causal events in the social world. So the failure of individualist theory to explain the law-like relations between tokens of social events is merely a realization of the expressive autonomy of social theory.

4.2.2. Kim's Notion of Strong Supervenience Applied to Social Explanation

Let me refine my claims about the ability of individualist theory to reduce our social ontology without being able to explain everything explained by social theory. In "The Mind-Body Problem: Taking Stock after Forty Years" Jaegwon Kim offers a reductive account of the mental based on his notion of strong supervenience. Kim suggests that his arguments my work in other domains, including the special sciences; in this section I will adapt them to show that they can work in social explanation.

According to Kim, higher level properties strongly supervene on lower level properties if and only if lower level indiscernibility entails higher level indiscernibility. So, for any social property or set of social properties S, there is some property or set of properties in the supervenience base I such that I is sufficient (though perhaps not necessary) for S. This understanding of supervenience does not entail (though the language indeed suggests) that the base properties are ontologically prior to or determinative of supervenient properties; at best it shows that the supervenient properties are strongly correlated with the base properties. In this paper Kim does not suggest any qualifications on the understanding of supervenience that would entail such priority, but he assumes that there are some.

The question faced by philosophers of mind is, 'how do you explain the supervenience relation between the mental and the physical?' Some potent answers include the identity theory, functionalism, emergentism, and perhaps every other physicalist theory of the mind. Taken this way, the question of the nature of supervenience is merely a restatement of the mind-body problem under the physicalist assumption that the mind is at base a physical entity. All physicalists assume some form of supervenience (broadly construed so as to include identity), so the question of how to explain supervenience just is the old fashioned question of the relationship between the mind and the body.

The physicalist world view often incorporates the view of the world as a set of levels, where different levels are distinguished by mereological relations: the entities at the lowest level are the composite parts of the entities at the next level, which are the composite parts of the entities at the next level, and so on. Quarks compose atoms, which compose molecules, which compose cells, which compose organisms, which compose social groups, and so on. There are properties that appear at each level that make that level distinct: electrical conductivity appears at the molecular level, metabolism appears at the cellular level, and rational choice appears at the level of intentional individuals. But this world view raises questions at each level: how does this level relate to the levels below it? How are biological properties related to physical and chemical properties? How is consciousness related to biological properties? How are social phenomena related to individual phenomena? We might try to give the same answer for every level-for example, by claiming that all levels ultimately reduce to particle physics. Or we might be reductionists about some levels but not others, or we

might be reductionists about some elements of a level but not others. Kim, for example, wants to be a reductionist about intentionality but not about qualia, both of which belong to the level of consciousness.

Nonetheless, we can assume that supervenience holds between each level. It follows that for any x and y in a level L, if x and y have all the same sub-L-level properties (Kim calls this microindiscernibility), they have all the same L-level properties. If we define microindiscernibility as demanding isomorphic decompositions of x and y into their proper parts, supervenience on the layered model turns out to be mereological supervenience. "A general claim of supervenience, then, becomes the Democritean doctrine that the world is the way it is because the microworld is the way it is" (175).

There are several important consequences of this view. Every object x with a property S has a unique decomposition into parts at the most fundamental level, and there is a set of lowest-level properties that belong to the parts of that decomposition. Furthermore, every object y with an isomorphic decomposition into fundamental parts that has all of the same lowest-level properties will also have S. This gives us ontological physicalism: the claim that every system (biological, mental, social) is entirely constituted by physical parts; and mereological supervenience: the claim that higher level properties (biological, mental, social) supervene on lower level properties. Social groups are the way they are because they are made of certain parts that stand in certain relations to one another. Of course, we have still not reduced the higher levels to the lower levels, because there is still the question of whether the supervenience

relation is explainable, or must be taken as primitive. Kim goes on to argue for the former.

First he defines 'second-order property,' suggesting that higher level properties are to be explained as second-order properties. A second-order property is a property F that ranges over a set of base properties B. F is the property of having one of the properties P in B where D(P). D specifies a condition on the members of B. For example, we can define 'being a primary colour' as a second-order property by letting B be the set of all colours and letting D contain only red, blue, and green. Having a primary colour is the second-order property of having one of the properties in B specified by D.

But the notion of second-order properties is not sufficient to do Kim's work: he must set some parameters on the sorts of conditions that D is allowed to set on B. So he defines 'functional property' as second-order properties over B where D involves a causal or nomic relation. "Functional properties are second-order properties defined in terms of causal/nomic relations among first-order properties" (177). Take the example of being a pain killer: both Aspirin and Tylenol have the property (causal power) of reducing pain in humans, but as a product of different chemical structures (ASA and acetaminophen). If social properties are functional second-order properties, then they are *extrinsic* or *relational* properties of the individuals that compose the groups. The chemical structure of Tylenol is an intrinsic property, but that that property numbs pain in humans is a property of the relation between Tylenol and nervous systems in humans. With this in mind we can hearken back to Kincaid's claim that the same event individualistically described (two men shaking hands) can count as many different

things (marrying their children, making a business agreement, declaring war), given different social contexts. Kim should agree: in each case we've changed the system that the property (hand-shaking) exists in, and since social properties are relational, we should expect changing the system that the relations operate in to change the effect of the relations.

Furthermore, given the requirement that D specifies a causal or nomic condition on B, we can expect that variation in the laws of nature, or in anything that can alter the causal powers of individuals, would alter the ability of individuals to realize social groups. But if we hold constant the relations between parts of a system of individuals and the laws that govern the causal powers of those individuals, we determine what social groups those individuals will be able to constitute. So the individuals are at most nomologically (not logically or metaphysically) sufficient for the social groups they constitute. So if a set of individualist properties (I1, I2,…,In) realize social properties (S1, S2,…Sn), then the S's are supervenient on the I's, but the supervenience is a nomological, not logical or metaphysical, necessity.

This view (call it 'individual realizationalism') explains the supervenience relation. The social supervenes on the individual because every social property is a second-order functional property with individual realizers. In systems of relations between individuals like the ones we find around here, having a certain social theoretic predicate is constituted by having a certain individual theoretic predicate. Having the individual predicate is sufficient for having the social predicate, and given that the second-order social predicate is only generated by quantification, having a social predicate is nothing substantive over and above having one of its individual realizers.

Instead of biconditional bridge-law correlations between social predicates and individual predicates (S↔I) we have identities (S=I).

Thus we have a reduction of the social to the individual. For each social property we offer a functional construal in terms of its causal or nomic relations to other properties: power (in certain contexts) is the ability to give orders and have them fulfilled, to punish those who don't obey, etc., just like heat is the ability to melt snow, cook meat, etc. We then go about looking for micro-level properties that fill the specified causal roles: in the case of heat, the mean kinetic energy of the molecules of a body; in the case of power, the conventional solution to a co-ordination problem about how to generate efficient behaviour in a particular situation. So, the Lewisian has a way to answer Kincaid. We can have reduction of the social to the individual without requiring that we can 'fully explain' social phenomena in individualist terms. As Kim notes, social theory is not eliminable. "From the ordinary epistemic and practical standpoint, second-order designators can be indispensable and unavoidable, and these designators introduce a set of useful and practically indispensable concepts that group first-order properties in ways that are essential for communicative purposes" (182). Second-order predicates let us talk about laws that relate more than individual tokens of social phenomena, and give 'full explanations' in Kincaid's sense, without committing us to any ontology beyond the ontology of individuals.

4.3.1. Gilbert's Notion of Collective Intentions

In this section we'll finally find a decisive reason to side with one of the two theories presented in the first two chapters of this thesis. It is drawn from one of the most influential non-reductive account of collective intentionality of the last several

years, offered by Margaret Gilbert in On Social Facts. The book has as its broadest goal a characterization of social groups. But along the way Gilbert offers an implicit account of collective intentionality. As a result, a sufficient understanding of her notion of collective intentionality demands a brief foray into her account of social groups.

To state Gilbert's point with almost unenlightening brevity, a social group can be characterized as some group of at least two people who have agreed to join forces toward some mutually accepted end. By 'joining forces' we agree to make our efforts as a group toward that end. In joining forces we constitute a *plural subject*, which means that each of us has put our own will into a 'pool of wills' (Gilbert borrows the term from Rousseau) that has a single goal (197-199). However, the acts that will be aimed at bringing about that end need not begin immediately. The plural subject need only be *jointly ready* to act when the appropriate time comes. Gilbert offers a technical characterization of joint readiness (197) that I need not go into here, as its technical sense is, for the purposes of this thesis, close enough to its colloquial sense.

We see Gilbert's understanding of collective intentionality in her talk about a pool of wills. This pool of wills is to be understood non-reductively. Gilbert distinguishes reductive from non-reductive accounts of collective intentionality via an influential claim made by Max Weber in sociology. Weber is famous for his slogan that any collective action must reduce to 'the actions of participating individual men' (quoted in Gilbert, 417). Both reductivists and non-reductivists should agree with this claim, but offer different explanations of what it implies. Both reductivists and non-reductivists would agree with Weber if he is only suggesting that our concepts about social groups imply that facts about the actions of such groups are constituted by facts

about the wilful acts of individual people. Gilbert calls this the *human intentional states* requirement. If this is all that Weber is proposing, then it should be amenable to everyone. It says merely that all human intentional states (including group intentional states) exist in the heads of individual humans. There is no Hegelian group mind floating above the minds of individuals.

The second interpretation of Weber distinguishes reductivists from non-reductivists. If Weber is interpreted as saying that social groups are constituted by the acts of what Gilbert calls 'singular agents' (a singular agent is an agent who acts to achieve its own goals), then the reductivist agrees and the non-reductivist does not. The non-reductivist is incensed by the assertion there are no collective acts that are not mere promotions of the goals of individuals. Gilbert thinks it is clear that the individualist is making a mistake here: the whole point of adding our will to a pool of wills is to achieve something that may not be our own end. Thus, for example, if the three members of the history department staff selection committee are trying to fill a vacant position from a shortlist of four applicants, each member might support one of the candidates. Thus each of the first three candidates is strongly supported by one committee member. But suppose that the fourth applicant was everyone's second choice. We might easily tell a story where the obvious rational choice for the group would be the fourth candidate, even though he was no one's preferred applicant. So, it seems obvious to Gilbert that there are collective actions that are not mere promotions of individual ends.

Gilbert agrees that if Lewis' account of convention is accurate, her goal cannot be met (318). If social conventions can be explained in the individualistic language of

Lewis' theory, there will be no explanatory role left for her 'plural subject.' So she launches four arguments against Lewis.

Lewis' theory rests heavily on game theory. His characterization is thus an idealization, since the game theory he relies on assumes perfect information and perfect rationality. In her first objection Gilbert asserts that the power of even such idealized rationality is not as great as we might hope. In particular, what I will call the 'conditionalization problem' shows that it is not necessary that ideally rational agents will follow precedents, choose salient options, or abide by conventions when they are faced with co-ordination problems.

The argument has its genesis with Richard Grandy, one of Lewis' reviewers. Grandy found a problem with the conditional preferences of co-ordinating agents. When faced with a co-ordination problem with a salient, precedented, or conventional solution, each player must reason thus. "I should choose the salient, precedented, or conventional solution, on the condition that the other players will. What will the other players do? I should try to replicate their reasoning. They will reason thus: 'I should choose the salient, precedented, or conventional solution, on the condition that the other players will. What will the other players do? I should try to replicate their reasoning. They will reason thus…' " This process could clearly be continued ad infinitum; in other words, it is a futile process. Precedents, salience, and conventions cannot help us because each co-ordination problem is a decision unto itself and in each such situation we must ask whether others will play their part in the salient, precedented, or conventional solution. Since we have no reason to believe one way or the other in this

particular case, salience, precedent, and convention cannot help us rationally to solve co-ordination problems (330-337).

Gilbert points out that induction over past cases of conformity will not benefit Lewis here (337). The stipulation that the players are rational agents presupposes that they have free will; free will blocks induction. Such agents act freely when reason does not determine their action, so we cannot assign probabilities to their acts in any this instance. The first time we solve a co-ordination problem by free action it will be common knowledge that we have done so, but this gives us no reason for ascribing a tendency toward a particular action for any of the players. Also the second time the group faces the same problem: since we have not established any tendencies for the players, we must assume that they will act freely; so we cannot rationally establish what they will do. Since this line of reasoning can be extended infinitely, she sees no reason to think that common knowledge of shared beliefs about tendencies to conform will insure that rational agents can solve co-ordination problems.

I think a rather brief response will suffice for this objection, even if we accept Gilbert's questionable assumptions about the behaviour of free agents[14]. Lewis' theory does rest on an idealization of games, and Gilbert has shown us something surprising about ideal agents. But this does not change the fact that actual agents in real co-ordination problems in fact use salience, precedent, and convention to solve co-ordination problems, and can count on others to do the same. Whether their actions are rational or not is beside the point; rational or not, they can use those mechanisms to generate social reality.

---

[14]We might wonder if we would have the same problem if we accepted a compatibilist or indeterminist view of free will. Gilbert doesn't seem to defend the theory behind her claims about the acts of free agents. I owe this point to Bruce Hunter.

Gilbert's second objection is that none of Lewis' conditions are individually necessary for having social conventions in a more ordinary sense of 'convention' (339-349). Her goal with this objection is at times extraneous to the discussion at hand; that his is no ordinary sense of 'convention' does not prevent Lewis' theory from adequately explaining the nature of social reality. So I will only review one of Gilbert's points where this inapplicability is at play; any unconsidered parts of the objection are subject to the same response. I'll focus on the parts of the objection that are a serious threat to Lewis' theory.

I'll first consider the one inapplicable argument. She suggests that we can have conventions on our normal understanding even when we do not face co-ordination problems (343-344). For example, it may be a convention that we sleep in beds while it is conventional in Java to sleep on mats. We do not need to be aiming at co-ordination in order to end up with this convention. It could be a product of other values of our society, such as comfort, consumption, and so on. But that we sleep in beds is not a bit of social reality, since we don't hold someone who sleeps elsewhere to be doing something wrong. So this is no counter-example to the aspect of Lewis' theory that I am interested in.

I'll move on to the elements of the objection that pose a threat to the aspect of Lewis' theory that I am interested in. She suggests that we may have conventions where there are no regularities in behaviour, as Lewis' theory requires (345-346). For example, imagine that it is a convention in our society that dinner guests should send thank-you notes to their hosts. But times change, people get busier, and some of the pre-conditions for the convention of note writing (such as attractive penmanship) are

weakened. People no longer regularly write notes. But even though there is no such regular behaviour, we might still say that it is a convention in our society that dinner guests write thank-you notes.

Gilbert leaves it to our intuitions to satisfy us that there is really a convention in this case. My intuitions are sharpened (in Gilbert's favour) if we add to the story that our Grandmothers say to us, "You know, you really ought to write your thank-you notes," and we all agree. In other words, we accept that x, a dinner party guest, counts as y, required to send a thank-you note.

It is not implausible to think that everyone in a situation like this would write their notes if it weren't for some excusing clause: "I'm always too busy" "No one can read my handwriting anyway" "I can't buy nice ink anymore" and so on. A hundred years ago we would not accept "I didn't get around to it" as a real excuse for not writing a note; now we do. Effectively, we no longer hold people responsible for writing notes; it has become acceptable in our society to not write notes. So what it really seems like is that we have a conflict in our conventions: it is conventional to write notes, but it is also conventional that you don't really need to write notes. The latter convention is stronger than the former, and since they are in conflict, the former becomes less and less regular (less and less of a convention) until it is no longer a convention at all. Often there is no longer any clear convention about what we ought to do, though we do have some intuitions about what conventions we ought to have-thus we reminisce about a bygone era where conventions were as conventions ideally are. Perhaps Gilbert's example hinges on the ambiguity between our intuitions about what

our conventions require of us and our intuitions about what our conventions would require of us in an ideal social world[15].

One of the Lewisian conditions on regularity R being a social institution in population P is that it is common knowledge in P that everyone in P expects everyone else to do R. Gilbert's response to this condition draws on the note-writing example above. First, she points out that Lewis uses 'expects' in the sense of 'believes it to be the case that everyone will do R,' not 'thinks it is incumbent on everyone to do R.' Then she points out that in the note-writing case there may not be common knowledge that note-writing is on the decline, but there may be some degree of uncertainty about how many people actually write notes. This uncertainty may lead many people to not expect others to conform, but this does not necessarily make us feel like there is no longer a note-writing convention. Individuals might think to themselves, "I really should write a note, but it's so seldom anyone does nowadays. Really, I'm sure no one else will." So, there appears to be the normative force of a convention at play, while there is no expectation that others will conform (346-349).

This objection is also covered by the response to the last objection. Gilbert's example seems dependant on a social situation where new conventions are arising, and indeed seem to be outweighing the old conventions (we now accept that 'I didn't get around to it' counts as an adequate excuse). In such a situation, it seems appropriate to hold that our expectations should be conditionalized: each individual expects most others would be conforming to the old convention if they weren't conforming to the new convention. Thus we might not be too surprised to hear from supporters of the old

[15] Thanks to Bruce Hunter for helping me with this last point.

conventions, 'kids these days have no respect. All caught up in their high-speed lifestyles, they have no time for the things that really matter in life.'

Finally, she argues that conventions need not be arbitrary in Lewis' sense (341). When deciding on a dress code for their meetings, a group might elect more formal attire so that people will not feel too relaxed. Then, that they wear formal attire is a convention, but it is a convention for a good reason. The solution is non-arbitrary.

It seems to me that this only shows that more formal attire is or ought to be highest on everyone's preference rankings. If one of the co-ordination equilibria in a co-ordination problem is preferred by all players then certainly it is rational to co-ordinate on that point, if possible. They are trying to decide on a dress code, so they are trying to co-ordinate their attire. Since they have to work to co-ordinate, there must be several options, all of which are better than non-co-ordination. It should not surprise us that they have a reason to select the best of those options.

Gilbert's third main criticism is that Lewis' conditions are not sufficient to capture the normative status of conventions (349-355). As Lewis agrees, there is a normative side of conventions-there is some sense in which we ought to follow conventions. But Gilbert points out that this is neither the ought of intrinsic value (Gilbert's expression) nor the ought of rationality (351-352). We see the moral sense of 'ought' when we say things like, 'you ought to treat others as ends-in-themselves because doing so is intrinsically valuable.' 'Ought' in this sense points to some purported intrinsic value of the action recommended. But this is at least not always the 'ought' of convention, given the arbitrary nature of (at least some of) our conventions.

It might be convention that we use red ink to write memos to one another, but it does not follow that anyone thinks that writing in red is intrinsically valuable (351).

Nor is the ought of convention the ought of reasons. Suppose it is a convention in a factory that new employees should not sit at tables with old employees to whom they have not been formally introduced. As a new employee approaches a pair of unfamiliar old employees seated for lunch, they look up and give a discouraging look. The new employee is then hailed by a formally introduced acquaintance, with whom he goes to sit. The acquaintance says to him, "You know, you really ought not to sit with people you haven't been introduced to. It goes against the conventions here." Clearly, 'ought' is being used here very differently from its use in the sentence, "You know, you really ought not to sit with people you haven't been introduced to. You know how the excitement upsets your digestion" (352).

How should we characterize the 'ought' of conventions? Gilbert points out that Lewis has no satisfactory answer (355). Indeed, Lewis seems to hold that the 'ought' of convention is the 'ought' of intrinsic value. Conforming to convention answers to our preferences and the preferences of those we are co-ordinating with, and answering to those preferences is something we ought to do because it is intrinsically valuable to do so.

But though Lewis holds that this value is intrinsic, he does not seem to hold that it is universal. He says that norms are regularities that we believe we ought to conform to. But, as Gilbert points out, the 'we' here refers to our society, the society of which Lewis, Gilbert, and I are a part, not whichever social group it is that holds the convention (353). This leaves open the possibility that there are societies that do not

hold that answering to our preferences and the preferences of those we are co-ordinating with is intrinsically valuable. So he has not shown that there is a conceptual link between conventions and any sense of 'ought.'

I'm not sure how a Lewisian should respond here. The most severe problem could be resolved by arguing that answering to our preferences and the preferences of those with whom we are co-ordinating is valuable across cultures, which seems reasonable to me, at least insofar as rationality is valuable. Rationality arguably has intrinsic value at least ceteris paribus. But this still leaves the problem that the ought of convention is not the ought of intrinsic value. Gilbert's argument is not conclusive, of course; just because the Lewisian theory has not yet captured the ought of convention does not show that Lewis' theory lacks the resources to do so. I will not press the point further here.

Lewis has another response to all of the foregoing criticisms. Gilbert seems to be taking an ordinary language approach to understanding 'convention': she is looking for places where Lewis' conditions are not met, but we still have some intuition that there is a convention at play. But Lewis is not engaged in such conceptual analysis; he is defining a technical term (which sounds like, and means something rather similar to, our ordinary language term 'convention'), but as he is coining a term that has its value in its ability to explain social phenomena, it is not a problem for him if its definition is not perfectly aligned with the ordinary language term it resembles. Lewis has promised to define no one's concept of 'convention' but his own.

Gilbert's fourth criticism is the most telling, as it offers a legitimate counter-example to Lewis' theory. She tells a story where all of Lewis' conditions are met, but

we do not have any social reality. Imagine this situation: it is common knowledge in P that: (1) everyone in P is a peculiar sort of conformist: they each prefer that if most members of a logical subset of members of P perform salient act A in recurrent situation S, and she is a member of that subset, then she does A also; (2) everyone reads the Daily News, which is generally reliable; and (3) yesterday's headline was, "in P most redheads clap their hands two times at sunrise."

Now, call the redheads in P 'H.' Given 1-3, it follows that it is common knowledge in H that: (4) most members of H expect that most members of H will clap two times at sunrise; (5) each member of H prefers to clap two times at sunrise given that most others do; (6) If most other members of H clapped 12 times at sunrise, each member of H would prefer to clap 12 times at sunrise; and (7) each member of H claps two times at sunrise (356-357). Here all of Lewis' conditions are met, but we would be hard put to say that there is a there is a collective intention for redheads in P to clap twice at sunrise. Lewis' conditions fail to require that the agents in an aggregate are co-operating. This is exactly the sort of counter-example Searle was looking for in the last chapter.

The Lewisian cannot appeal to the conceptual analysis/theoretical definition distinction to save herself here. Lewis' term of art was 'convention', but the problem here is that, if the reduction of collective intentionality was sufficient, we would have a collective intention, and clearly we do not. The Lewisian theory does not adequately capture the phenomenon we want explained.

Clearly, a defence of individualism requires an adjustment to the definition of co-ordination so as to exclude this case while using only individualistic idiom. So, for

example, Lewis can't say that individuals must be co-ordinating so as to achieve some goal that can be only accomplished by people intending to work together, since 'working together' may not belong to the individualist's vocabulary. I am concerned primarily with Gilbert's reliance on conformism in generating this example. There seems to be a significant violation of the spirit of Lewis' account. Methodologically, he proceeds by offering several paradigm examples of co-ordination problems, and arrives at his characterization of co-ordination problems by attempting to capture what is held in common by all of the paradigm examples. But the case of conformance is importantly disanalogous with all of those paradigms. In Gilbert's example, we have more than one proper co-ordination equilibrium. But the value derived from each of those co-ordination equilibria lies solely in the fact that they are co-ordination equilibria. Think back to chapter two, and the seven examples I give of co-ordination problems (all of them are versions of Lewis'; he offers 11 examples). Lewis' co-ordinators want to co-ordinate so that they can ski together, so that they can find their friend, so that they can survive, and so on. Gilbert's co-ordinators want to co-ordinate just so that they can co-ordinate. But all of the paradigm examples suggest that co-ordination should be a means-to-an-end, not an end-in-itself. If this move is ad hoc, it may not be a dangerous species: co-ordination for its own sake really should have been excluded from Lewis' original account, since he's trying to capture the sense in which we organize ourselves so that we can accomplish something that we would otherwise be unable to accomplish. We are not organizing just for the sake of organizing ourselves. It does not help Gilbert's cause to point out that organizing ourselves is something we couldn't accomplish if we didn't organize ourselves. This is a tautology, and when we

eliminate the redundancy we find a substantive difference between real co-ordination problems and instances of conformism. When we co-ordinate we say, 'let's align our actions so that we can do *y*,' not simply, 'let's align our actions.'

But this response will not work. There are some cases that are properly called conventions, but we are only co-ordinating for co-ordination's sake. In the case mentioned above where memos in an office are written only in red ink, the goal may be nothing more than conformity, or the sense of neatness and order that conformity provides. Perhaps company policy demands merely that memos be written on the 'memo' form and have a particular sort of communicative intent. But a particular department might decide to impose a further condition on memo writers. The electrical engineers, for example, might determine that x, a person who is writing a memo (on memo paper, with the right sort of intent), counts as y, a person who should be writing in red ink. Here we have a status-function, but the goal of its imposition might be merely keeping those pesky electrical engineers from expressing their individuality. Furthermore, the electrical engineers might have determined for themselves that their individuality needs to be curbed, and so instituted this status-function. But now, like in Gilbert's counter-example, we have conformism for the sake of conformism. Finally we have a decisive reason to choose a side in the debate between Searle and Lewis. Even if the Lewisian responses to Gilbert's first three criticisms are adequate, he certainly has as yet no adequate response to her counter-example. The burden of proof is returned to Lewis, and it looks like a serious one.

4.4.1. Gilbert on Searle

Gilbert would not be satisfied with this conclusion. She holds that Searle's position is not anti-reductionist enough[16]. She contends that he does not truly take collective intentionality as primitive. Instead of reducing collective intentionality to a set of individual intentions (augmented by beliefs, solutions to co-ordination problems, or whatever), he reduces them to another sort of individual mental state. As she writes, for Searle, "For *them* to intend is for *each of them* to think a particular thing" (Sociality and Responsibility, 157). This is not sufficiently robust to support the plural subject Gilbert takes to be necessary for any social phenomenon. In particular, it does not entail the obligations of each agent to carry out the joint action specified by the collective intention.

Let's take an example from On Social Facts. Suppose we form a poetry reading group that intends weekly to analyze works of a given author. We might begin thus: on our first week, we have chosen to analyze two poems by a given author. Our method in analysis is thus: we first read one poem and arrive at an analysis; we then use that interpretation to inform our interpretation and appreciation of the second. Suppose, for whatever political reasons, we collectively agree that the rather bathetic last line of the first poem is truly moving. We then move on to the second poem, and at one point in the discussion someone says, "This line is much more insightful than the bathetic conclusion of the last poem." Some members of the group might rightfully find this comment rather incendiary, as it violates the prior agreement (289). If there are such

[16] Similar charges have been laid against Gilbert. In "Doing Things with Others" Annette Baier argues that Gilbert's self-styled 'weak analytic individualism', which allows that the concept of individuality does not presuppose collectivity concepts, cannot adequately explain the nature of reasoning. Baier argues that private reasoning is parasitic on public argumentation. Modus ponens, she suggests, can only justify moves in reasoning in a community where modus ponens is used to justify moves in reasoning.

sentiments, they should have been voiced earlier. Now is not the appropriate time. Once the members of a group have individually put their will into the pool of wills that intends to accept that a bathetic phrase is moving, they are obligated to accept that the phrase is moving and to act in accordance with their acceptance, whether they personally agree that the phrase is moving or not.

Gilbert suggests that a group with merely a Searlean collective intention may not have any such obligation. A Searlean collective intention is an aggregate of parts-the (collective) intentions of each member of the collective. Gilbert prefers her notion of a 'joint commitment', which is a single, unified, indivisible whole, rather than an aggregate. Agents together make a single joint commitment, not a series of individual commitments, to act in accord with the group intention. On her understanding (not on Searle's), no party can "rescind the joint commitment unilaterally. Nor can they rescind any "part" of it, since it does not have any parts. It binds them together in the service of the cause in question until they concur in dissolving the bond they have created" (Sociality and Responsibility 158). As a Searlean collective intention is neatly divided into constitutive parts, we should expect at least the possibility of unilateral rescission, and Searle has given us no reason to think that such unilateral defection would be morally blameworthy.

In formulating a response on Searle's behalf I'll borrow from Michael Bratman's "Shared Intention". Gilbert gives Bratman the same criticism as she gives Searle. But Bratman thinks that it is a good thing if our account of collective intentionality (shared intention in his terminology) does not imply that we have obligations to act in accord with the collective intention. He offers this counterexample:

suppose we have two singers who enjoy singing duets together. They might plan together to sing a duet on a particular date, yet be perfectly clear that they may decide to withdraw their intention at any point. Though we have a collective intention to sing together, they do not have a moral obligation to act in keeping with their collective goals. So, Bratman concludes that our account of collective intentionality ought not to imply that we have moral obligations to stick with our collective goals. We can at times unilaterally rescind collective intentions, so it is a shortcoming of Gilbert's position that she precludes this possibility.

4.5.1. Conclusion

In this thesis we've seen two very distinctive accounts of the construction of social reality. Closer examination, however, proved that the two share many common features on the most important questions about how we get a social world. Both are able to recognize the centrality of conventions in the world of social institutions. The social world, as Searle taxonomizes it, is a polymorphous realm, and conventions are not at play in all facets of that realm. But in the domain of status functions, conventions rule. So argues Lewis, and though Searle argues for the value of rules over convention, his understanding of 'rule' tacitly supports Lewis' position. Thus we find a path of inquiry that may answer the question posed in the opening sentences of this thesis: perhaps weddings, parking tickets, undergraduate degrees and their kin are distinguished from molehills, electrons, and gravity by the role of convention in their generation. The central point of difference between the two theorists we've focused on was over the ontological status of collective intentionality, though both views recognize the centrality of collective intentionality to an account of status-functions. Collective

intentions exist without conventions, but the converse is not true. There are social phenomena, even co-ordination problems, that exist in the absence of conventions. But there are no conventions that exist if there are no social phenomena. Convention is a sufficient but not necessary condition for sociality. The role of our whimsical conventions is explaining the steadfast evanescence of our status-functions.

We saw a strong defence, but the ultimate failure, of the Lewisian approach to the reduction of collective intentionality. This, of course, does not close the issue. Kim's approach, for example, does not hinge on the details of Lewis' account, and we might wonder if Gilbert's arguments apply to Kim's reductionism. But that is another question. The project at hand is decided in Searle's favour.

Bibliography

Alanen, Lilli, Sara Heinämaa, and Thomas Wallgren, eds. Commonality and Particularity in Ethics. New York: St. Martin's Press, 1997.

Baier, Annette. "Doing Things with Others: The Mental Commons" in Commonality and Particularity in Ethics, ed. Lilli Alanen, Sara Heinämaa and Thomas Wallgren. New York: St. Martin's Press, 1997.

Bratman, Michael. "Shared Intention" in Faces of Intention. Cambridge: Cambridge University Press, 1999.

Cohen, Phillip, Jerry Morgan, and Martha Pollack, eds. Intentions in Communication. Cambridge, Mass.: MIT 1990.

Cooney, Brian, ed. The Place of Mind. Scarborough: Wadsworth, 2000.

Davidson, Donald. "The Method of Truth in Metaphysics" in Inquiries into Truth and Interpretation. Oxford: Clarendon Press, 1984.

Dretske, Fred. "If You Can't Make One, You Don't Know How it Works" in The Place of Mind, ed. Brian Cooney. Scarborough: Wadsworth, 2000.

Gilbert, Margaret. On Social Facts. Princeton: Princeton U.P., 1992.

Gilbert, Margaret. Sociality and Responsibility. Lanham: Rowman and Littlefield, 2000.

Hellman, G., and F. Thompson. "Physicalism: Ontology, Determination, and Reduction"
in The Journal of Philosophy 72, 1975.

Kim, Jaegwon. "The Mind-Body Problem: Taking Stock After 40 Years" in The Place of Mind, ed. Brian Cooney. Scarborough: Wadsworth, 2000.

Kincaid, Harold. "Reduction, Explanation, and Individualism" in Readings in the Philosophy of Social Science, ed. Michael Martin and Lee C. McIntyre. Cambridge, Mass.: MIT, 1995.

Levine, Joseph. "Recent Work on Consciousness" in American Philosophical Quarterly vol. 34 no. 4 October 1997.

Lewis, David. Convention: A Philosophical Study. Cambridge, Mass.: Harvard, 1969.

MacDonald, Graham and Philip Pettit. Semantics and Social Science. Boston: Routledge and Kegan Paul, 1981.

Martin, Michael and Lee McIntyre, eds. Readings in the Philosophy of Social Science. Cambridge, Mass.: MIT, 1995.

Millikan, Ruth. "Language Conventions Made Simple" The Journal of Philosophy 95, 1998.

Searle, John. Intentionality: An Essay in the Philosophy of Mind. Cambridge: Cambridge UP, 1983.

Searle, John. "Collective Intentions and Actions" in Intentions in Communication, ed. Cohen, Morgan, and Pollack. Cambridge, Mass.: MIT, 1990.

Searle, John. The Rediscovery of the Mind. Cambridge, Mass.: MIT, 1992.

Searle, John. The Construction of Social Reality. New York: Basic Books, 1995.